



Publikationen der Deutschen  
Gesellschaft für Photogrammetrie,  
Fernerkundung und Geoinformation e.V.

---

**Band 31**

**2023**



**Beiträge**

**43. Wissenschaftlich-Technische  
Jahrestagung der DGPF**

22.-23. März 2023  
in München

***DGPF-Jahrestagung  
Photogrammetrie - Fernerkundung - Geoinformation - 2023***

ISSN 0942-2870

DOI: 10.24407/KXP:184104914X

Thomas P. Kersten & Nora Tilly, Hrsg.

Publikationen  
der Deutschen Gesellschaft für  
Photogrammetrie, Fernerkundung  
und Geoinformation e.V.



---

**Band 31**

**2023**

**Beiträge**

43. Wissenschaftlich-Technische  
Jahrestagung der DGPF

22.-23. März 2023  
in München

ISSN 0942-2870

Thomas P. Kersten & Nora Tilly, Hrsg.

DOI: 10.24407/KXP:184104914X

ISSN 0942-2870

Publikationen der  
Deutschen Gesellschaft für Photogrammetrie, Fernerkundung  
und Geoinformation (DGPF) e.V.  
Band 31, 401 S., Hamburg und Köln 2023  
Hrsg.: Thomas P. Kersten & Nora Tilly

© Deutsche Gesellschaft für Photogrammetrie, Fernerkundung  
und Geoinformation (DGPF) e.V.  
Stuttgart 2023

Geschäftsstelle der DGPF  
c/o Universität Stuttgart  
Institut für Photogrammetrie (ifp)  
Geschwister-Scholl-Str. 24D  
70174 Stuttgart  
E-Mail: [geschaeftsstelle@dgpf.de](mailto:geschaeftsstelle@dgpf.de)

Redaktion:

Thomas P. Kersten  
HafenCity Universität Hamburg  
Labor für Photogrammetrie & Laserscanning  
Henning-Voscherau-Platz 1, 20457 Hamburg  
E-Mail: [Thomas.Kersten@hcu-hamburg.de](mailto:Thomas.Kersten@hcu-hamburg.de)

Nora Tilly  
Stadt Köln - Amt für Stadtentwicklung und Statistik  
Willy-Brandt-Platz 2, 50475 Köln  
E-Mail: [Nora.Tilly@uni-koeln.de](mailto:Nora.Tilly@uni-koeln.de)

# INHALTSVERZEICHNIS

## KEYNOTE

GOTTFRIED MANDLBURGER & PATRICK WESTFELD: Ein Überblick über optische Methoden in der Hydrographie. ....	1
---	---

## BEITRÄGE KARL-KRAUS-NACHWUCHSPREIS

JONAS MEYER: AI-Based 3D-Detection of Parked Vehicles on a Mobile Mapping Platform using Edge Computing. ....	22
CHRISTOPH BAUMANN, MOSTAFA KIANI SHAHVANDI, ALEXANDRA DUCKSTEIN, MALTE WESTERHAUS & BENEDIKT SOJA: Using Machine Learning to Predict Earth Deformation from InSAR Time Series. ....	30
ADRIAN WENG, FRANZISKA WILD & EBERHARD GÜLCH: Evaluation of Historical Aerial Images in a Landslide Area. ....	36
NINO LENZ & MICHAEL GREZA: Simulation of Earth Observation Data Utilizing a Virtual Satellite Camera. ...	50
NIELS DEDRING, VALERIE GRAW & ANDREAS RIENOW: HEC-RAS-Simulation einer Glacial Lake Outburst Flood zur Ermittlung von möglichen Auswirkungen am Beispiel des Chamlang North Tsho (Lake 464) im Hongu Valley, Nepal. ....	61
BINGXIN KE, CORINNE STUCKER & KONRAD SCHINDLER: Spot the Difference: Learned DSM Updating. ....	69
HANNES NÜBEL & PATRICK TUTZAUER: Improving True Ortho Quality by Removing Moving Vehicles. ....	78

## **KURZVORTRÄGE 1 (POSTER)**

JIWEI HOU, MONA GOEBEL & DOROTA IWASZCZUK: Real-Time Tracking and 3D Dense Reconstruction Based on ORB-SLAM3 Extensions using a Depth Camera. ....	87
THEA MINNICH & ANSGAR BRUNN: Mobile 3D-Vermessung zur Dokumentation von Gasleitungen. ....	98
DANIEL HEIN, MATTHIAS GESSNER, THOMAS KRAFT, JULIA GONSCHOREK, HENRY MEIßNER, TILMAN BUCHER & RALF BERGER: Real-time Distribution of an Airborne Situational Picture into Command and Control Systems. ....	110

## **KURZVORTRÄGE 2 (POSTER)**

ALEXANDER P. WALMSLEY, MAREN LINDSTAEDT, MIKE SCHNELLE & THOMAS P. KERSTEN: Virtual Reality for Spatial Research in Archaeology: Prototyping 3D Reconstruction Tools for the Site of Yeha, Ethiopia. ....	123
NINA KRUEGER & STEPHAN PLABST: An Approach to Develop a Forestry Data Space Using Geoinformation and Remote Sensing Data for Forest Management and Protection. ....	134
MAXIMILIAN HELL, MELANIE BRANDMEIER & ANDREAS NÜCHTER: Transferability of Deep Learning Models for Land Use/Land Cover Classification. ....	142
SAJA SALATTNA, JULIANE BENDIG, CHRISTINA KUCHENDORF, CHRISTOPH JEDMOWSKI & UWE RASCHER: Monitoring Crop Seasonal Development under Biochar Treatment with Unmanned Aerial Vehicle (UAV)-based Multispectral Imagery. ....	150
DAVID COLLMAR, VOLKER WALTER, MICHAEL KÖLLE & UWE SÖRGEL: A Two-Step Approach for the Acquisition of Individual Tree Outlines using Paid Crowdsourcing. ....	163

## **SENSORIK**

- MICHAEL GREZA, LUDWIG HOEGNER, PHILIPP-ROMAN HIRT, ROBERT ROSCHLAUB  
& UWE STILLA:  
Satellite Network Bavaria – Mission and Data Processing. .... 174
- FLORIAN PÖPPL, MARTIN PFENNIGBAUER, ANDREAS ULLRICH,  
GOTTFRIED MANDLBURGER, HANS NEUNER & NORBERT PFEIFER:  
Modelling of GNSS Positioning Errors  
in a GNSS/INS/LiDAR-integrated Georeferencing. .... 183

## **FORST UND AGRAR**

- SAMUEL WILDHABER, LUKAS V. GRAF & HELGE AASEN:  
Assessing the Potential of High-Resolution Satellite Constellations  
for Agricultural Monitoring. .... 197
- CHRISTOPH HÜTT:  
Flying Parameters Experiment of a UAV LiDAR  
over a Winter Wheat Field-trial. .... 206
- SEBASTIAN BRIECHLE:  
Instanzsegmentierung von stehendem Totholz  
mittels Mask R-CNN und CIR-Bildern. .... 213
- VIKTORIJA HNATUSHENKO, VOLODYMYR HNATUSHENKO, VITA KASHTAN  
& CHRISTIAN HEIPKE:  
Detection of Forest Fire Consequences on Satellite Images  
using a Neural Network. .... 223

## **OPTISCHE 3D-MESSTECHNIK**

- ANSGAR GREIWE, VOLKER SPRECKELS, ANDREAS SCHLIENKAMP,  
RAINER BRECHTKEN & SIMON DAMBERG:  
Qualitätsuntersuchungen UAV-getragener Kamerasysteme  
am Testfeld Bochum. .... 233
- JÜRGEN SELENSKI, OLIVER REUß, JONAS ROTH, VERENA HÄNDLER, TOMASZ BADECKI,  
UWE SCHIRMER, MARKUS BOGNER, ANDREAS WAGNER & ANSGAR BRUNN:  
Integration von 3D-Laserscanning in die Rohstoffgewinnung  
im Untertagebau. .... 251

## **RADAR**

- BASTIAN GESSLER, SILKE MECHERNICH & BJÖRN BASCHEK:  
Detektion großräumiger Niedrigwassersituationen durch satelliten-  
basierte Wasser-Land-Grenzen-Bestimmung („Sat Land-Fluss“). ..... 263
- GWENDOLYN DASSER, JESSICA MUNCH, YVES BÜHLER, PERRY BARTELT  
& ANDREA MANCONI:  
Applied Space-borne Remote Sensing to Identify Mass Movements  
and the Exemplary Modelling of Potentially Catastrophic Failures  
in the Bhagirathi Area, India. .... 276

## **UMWELTANALYSE**

- KATRIN GÖLZ, SYLVIA HOCHSTUHL & ANTJE THIELE:  
Silbermöwe oder Heringsmöwe? -  
Klassifikation von Seevögeln in UAS-Aufnahmen. .... 288
- LINA E. BUDDE, DOROTA IWASZCZUK, LAYTH SAHIB & JÜRGEN BORN:  
Ko-Registrierung von Sentinel-1 und Sentinel-2 Zeitreihen  
für die Detektion von Trinkwasserleckagen: ein Methodenvergleich. .... 301

## **OPTISCHE 3D-MESSTECHNIK**

- JAKOB SCHMIDT, ANDREAS EICHHORN & DOROTA IWASZCZUK:  
Punkt- und ebenenbasierte Detektion von Ecken und Kanten  
in Innenraum-Punktwolken. .... 310

## **BATHYMETRIE**

- MICHAEL GRÖMER & GOTTFRIED MANDLBURGER:  
Vergleich verschiedener Unterwasser-Photogrammetrie-  
Inspektionsmethoden zur Bewertung von unterwasserliegenden  
Stahlwasserbauwerken. .... 322
- MICHAEL BLEIER, CHRISTOPH MUNKELT, MATTHIAS HEINZE,  
CHRISTIAN BRÄUER-BURCHARDT, STEFAN MARX, DIRK FROMME,  
CHRISTIAN DETTMER, HELGE ANDREAS LAUTERBACH & ANDREAS NÜCHTER:  
UWSensor: Ein optisches Unterwasser-3D-Scannersystem  
für mobile Kartierung. .... 332

## **VIRTUAL UND AUGMENTED REALITY**

HELGE OLBERDING:

Visualisierungs- und Filterungsmethoden von CityGML-Daten  
in einer VR-Umgebung. .... 347

FERDINAND MAIWALD, CLEMENS BECK & SANDER MÜNSTER:

Jena4D – Ein virtuelles Stadtgeschichtsbuch zum Mitmachen. .... 358

PETER BAUER & WERNER LIENHART:

Experiences from the Creation of a Virtual Tour  
through a Stalactite Cave with VR Gear. .... 368

THOMAS P. KERSTEN, DANIEL DRENKHAN & SIMON DEGGIM:

Analyse der Echtzeit-Visualisierung für Virtual Reality-Anwendungen  
am Beispiel der Festung Al Zubarah in Katar. .... 381

**AUTORENVERZEICHNIS** ..... 397

**PROGRAMMKOMITEE** ..... 401



# Ein Überblick über optische Methoden in der Hydrographie

GOTTFRIED MANDLBURGER<sup>1</sup> & PATRICK WESTFELD<sup>2</sup>

*Zusammenfassung: Die Geschichte des Einsatzes optischer Methoden für hydrographische Anwendungen reicht bis in die Mitte des letzten Jahrhunderts zurück. Seit diesen Anfängen haben sich sowohl die passiven als auch die aktiven Techniken erheblich weiterentwickelt. Mit immer neuen Sensoren und Plattformen, die zur Verfügung stehen, gibt es einen kontinuierlichen wissenschaftlichen Fortschritt im Bereich der Erfassung der Unterwassertopographie seichter Gewässer. Dieser Artikel gibt einen Überblick über die verwendeten optischen Methoden, die in passive und aktive Techniken unterteilt werden können. Zu den passiven Methoden gehören die spektrale Tiefenbestimmung sowie die Mehrmedienphotogrammetrie, die aktive Methode ist die Laserbathymetrie. Ein weiterer Ansatz zur Kategorisierung optischer Methoden basiert auf der verwendeten Trägerplattform, wobei Satelliten, bemannte Flugzeuge, unbemannte Flugsysteme und neuerdings auch ferngesteuerte Unterwasserfahrzeuge für die Kartierung von Gewässern in unterschiedlichen Maßstäben, Auflösungen und Genauigkeiten eingesetzt werden. Eine dritte Möglichkeit der Klassifizierung besteht darin, ob sich der Sensor über oder unter dem Wasserspiegel befindet. Der Beitrag beleuchtet auch die wichtigsten Datenverarbeitungstechniken, im Besonderen (i) die Refraktionskorrektur, (ii) die Signalverarbeitung des radiometrischen Inhalts und (iii) die Analyse des rückgestreuten Wellenformsignals in der Laserbathymetrie. In jüngster Zeit spielen Techniken des maschinellen Lernens nicht nur bei der Objektklassifizierung, sondern auch bei der regressionsbasierten Tiefenschätzung eine immer wichtigere Rolle. Abschließend wird auf die vielfältigen Anwendungsmöglichkeiten eingegangen, die von der Kartierung schiffbarer Kanäle für die sichere Schiffsnavigation über die Katastrophendokumentation, die Hochwassersimulation, den Küstenschutz, die Modellierung von ufernahen Lebensräumen, die Restrukturierung von Flüssen, die Überwachung hydromorphologischer Veränderungen, die Wartung von Wasserkraftwerken, die Kartierung von Offshore-Infrastruktur, die Kartierung von Korallenriffen bis hin zur Unterwasserarchäologie reichen. Dem Beitrag ist ein Vorwort von Patrick Westfeld, dem Leiter des Forschungs- und Entwicklungsbereichs des Referats Geodäsie, Seevermessung und Wracksuche am Bundesamt für Seeschifffahrt und Hydrographie (BSH), vorangestellt.*

## Präambel

Das vorliegende Manuskript stellt eine Kurzfassung eines open-access Überblicksartikels über optische Methoden in der Hydrographie dar, der Ende 2022 im International Hydrographic Review (IHR) erschienen ist (MANDLBURGER 2022). In der vorliegenden, gekürzten Form stellt der Beitrag

---

<sup>1</sup> TU Wien, Department für Geodäsie und Geoinformation, A-1040 Wien, Österreich,  
E-Mail: gottfried.mandlburger@geo.tuwien.ac.at

<sup>2</sup> Bundesamt für Seeschifffahrt und Hydrographie, Referat Geodäsie, Seevermessung und Wracksuche,  
D-18057 Rostock, Deutschland, E-Mail: patrick.westfeld@bsh.de

den Rahmen für die Bathymetrie-Plenarsitzung bei der 43. wissenschaftlich-technische Jahrestagung der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation in München dar. Diese Sitzung besteht aus einem eingeladenen Vortrag und einer anschließenden moderierten Podiumsdiskussion mit Expert:innen aus dem Bereich Wissenschaft, Industrie und Verwaltung. Der eingeladene Vortrag wird von Dr.-Ing. Patrick Westfeld, dem Leiter des Forschungs- und Entwicklungsbereichs des Referats Geodäsie, Seevermessung und Wracksuche am Bundesamt für Seeschifffahrt und Hydrographie (BSH), gehalten. Von ihm stammt auch das Vorwort zu diesem Beitrag. Patrick Westfeld ist Chefredakteur des IHR, dem offiziellen Publikationsorgan der International Hydrographischen Organisation (IHO).

Die IHO ist eine zwischenstaatliche Organisation, die sich dafür einsetzt, dass alle Meere, Ozeane und schiffbaren Gewässer der Welt vermessen und kartiert werden, um insbesondere die sichere Navigation in der Schifffahrt zu gewährleisten, das Küstenzonenmanagement zu fördern, die maritime Raumordnung zu ermöglichen und die Meeresumwelt nachhaltig zu schützen.

## Vorwort

Die Hydrographie liefert die Grundlagen für die nachhaltige wirtschaftliche Nutzung und den Schutz unserer Meere und Binnengewässer mit ihren Küsten und Ufern. Sie schafft die Voraussetzungen dafür, dass Schiffe sicher und effizient fahren, dass Fischerei möglich ist, dass Windparks gebaut und Kabel gelegt werden können. Hydrographische Daten sind ebenso wichtig für präzise Wasserstandsvorhersagen und den Schutz der Küsten- und Uferbereiche mit ihren Bewohnerinnen und Bewohnern. Auch sind ihre Informationen eine unverzichtbare Basis für politische und fachliche Entscheidungen für den Gewässerschutz und die damit in Zusammenhang stehende Entwicklung von Zukunftsszenarien. Das vorrangige Ziel der Hydrographie ist somit, das Wissen über unsere Gewässer zu erweitern, um sie verantwortungsvoll und sicher nutzen zu können und als Lebensraum zu schützen. Die Hydrographie ist der Schlüssel zu unseren Gewässern!

Die Hydrographie als ein Teilgebiet der Geodäsie ist eine Fachdisziplin, die im Hintergrund arbeitet und die nur wenige Fachleute kennen. Selbst für die Geodätinnen und Geodäten ist sie vielfach fremd. Deren besondere, ja strategische Bedeutung bleibt erst recht verborgen. Der Nutzen tritt eben selten unmittelbar zu Tage. Lassen Sie uns gemeinsam die gesellschaftliche und politische Aufmerksamkeit der durch die Vereinten Nationen für die Jahre 2021 bis 2030 ausgerufenen Internationalen Dekade der Meeresforschung für Nachhaltige Entwicklung (UN Ozeandekade) nutzen, um die Hydrographie und die benachbarten Fachdisziplinen Photogrammetrie, Fernerkundung und Geoinformation in den Fokus der Öffentlichkeit zu rücken. Lassen Sie uns gemeinsam zeigen, dass durch unsere interdisziplinäre Expertise die Gewässerumwelt genau und zuverlässig vermessen, fundiert analysiert und themenspezifisch visualisiert werden kann.

„The science we need for the ocean we want“ lautet das Motto der UN Ozeandekade, einer globalen Kampagne mit dem ambitionierten Ziel, wissenschaftsbasierte Lösungen für den Schutz und die nachhaltige Nutzung der Meere zu entwickeln und umzusetzen. Peter Thomson, der Sonderbeauftragte des Generalsekretärs der Vereinten Nationen für die Ozeane, erweiterte 2021 während einer Rede zu den Hundertjahrfeierlichkeiten der Internationalen Hydrographischen Organisation (IHO) dieses Motto: „The data we require, for the science we need, for the ocean we want.“ Die

durch bathymetrische Vermessung erhobenen Geobasisdaten der topografischen Gestalt von Gewässerbetten und Meeresböden bilden die notwendige Grundlage für raumbezogene Fachdaten. Sie sind folglich von fundamentaler Bedeutung, um die durch die UN Ozeandekade identifizierten Herausforderungen zu adressieren und die gesetzten Ziele zu erreichen.

Gleichzeitig betonte Thomson, wie wichtig eine enge Zusammenarbeit ist, um unsere Gewässer nachhaltig zu schützen. Im folgenden Übersichtsartikel von Gottfried Mandlbürger (TU Wien) wird auf hervorragende Weise deutlich, wie unterschiedliche Methoden, Verfahren und Systeme aus Photogrammetrie und Fernerkundung einen hydrographischen Beitrag weit über die klassische Gewässervermessung mit Echoloten hinaus leisten. Zum Einsatz kommen können zum Beispiel zur Hydroakustik komplementäre Verfahren wie Airborne Laserscanning Bathymetrie und multispektrale Satellitenbilddatenauswertung. Auch findet Gewässervermessung heutzutage nicht ausschließlich schiffsgestützt statt: Neben Flugzeugen und Satelliten haben unbemannt fliegende, schwimmende und tauchende Drohnen längst Einzug als Sensorträger gehalten (Abb. 1). Mandlbürger's Beitrag zeigt die vielfältigen Facetten der Hydrographie und ist ein Beleg für eine gelungene geowissenschaftliche und ingenieurpraktische Zusammenarbeit.

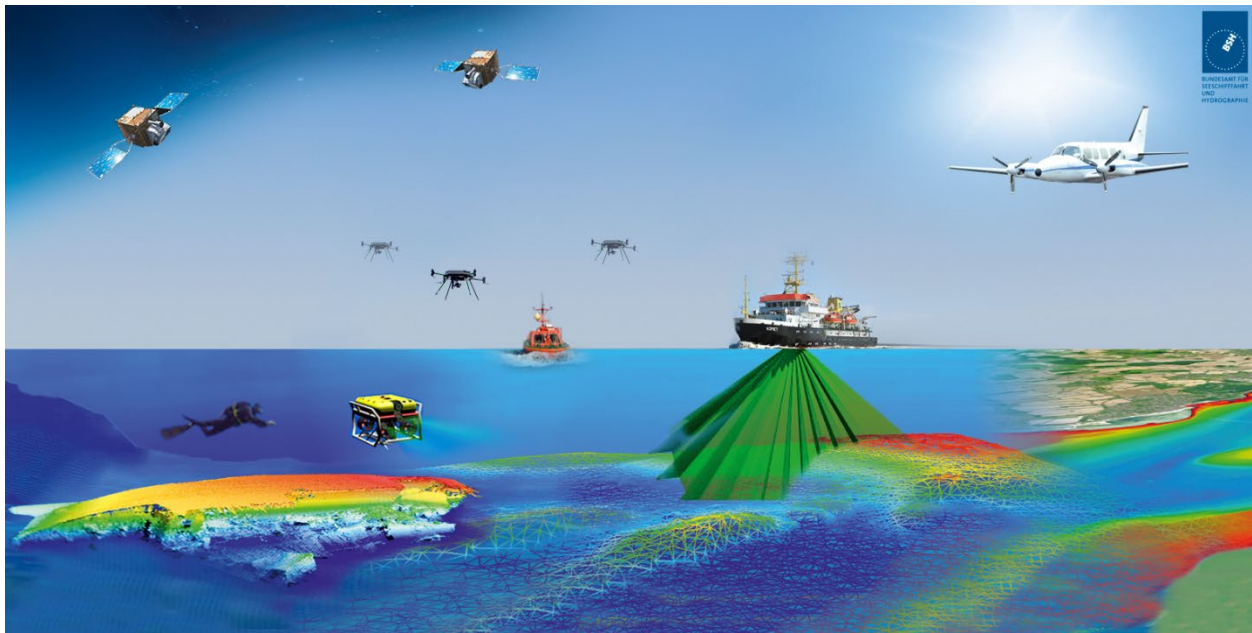


Abb. 1: Hydrographie – Messen mit allen Sinnen: Komplementäre Messverfahren und Sensorträgerplattformen zur bathymetrischen Datenerfassung.

## 1 Einleitung

Nach einer Definition der Internationalen Hydrographischen Organisation (IHO) ist die Hydrographie der Zweig der angewandten Wissenschaften, der sich mit der Messung und Beschreibung der physikalischen Eigenschaften von Ozeanen, Meeren, Küstengebieten, Seen und Flüssen sowie mit der Vorhersage ihrer zeitlichen Entwicklung befasst, und zwar zum vorrangigen Zweck der Sicherheit der Schifffahrt und zur Unterstützung aller anderen maritimen Aktivitäten, einschließlich

der wirtschaftlichen Entwicklung, der Sicherheit und Verteidigung, der wissenschaftlichen Forschung und des Umweltschutzes (IHO 2022). In diesem Zusammenhang wird auch der Begriff Bathymetrie verwendet. Beide Wörter stammen aus dem Griechischen, und während Hydrographie übersetzt „Wasser beschreiben“ bedeutet, bedeutet Bathymetrie „Tiefenmessung“. Bathymetrie ist also ein spezifischerer Begriff, aber beide können austauschbar im Sinne der Kartierung der Unterwassergeometrie verwendet werden.

Die genaue Kenntnis der Form und Veränderung der Unterwassertopographie und der Objekte ist die Grundlage für eine Vielzahl von sozioökonomischen und ökologischen Themen. Zu den ersten gehören die Sicherheit der Schifffahrt, die Bewertung des Hochwasserrisikos, die Planung von Gefahrenzonen und Schutzmaßnahmen. Zu den letzteren zählen die Wiederherstellung von Küsten- und Schwemmlandgebieten, die Überwachung ihres Zustands und ihrer Veränderungen sowie die Bewertung der jeweiligen Auswirkungen auf die aquatischen Lebensräume (Hydrobiologie, Lebensraummodellierung auf Mikro- und Mesoebene usw.).

Optische Methoden eignen sich gut für die Erfassung der Bathymetrie von klaren und flachen Küsten- und Binnengewässern mit Tiefen  $< 60$  m aus der Luft, sind aber für tiefere Gewässer aufgrund der hohen Lichtabsorption im Wasser ungeeignet. Befindet sich die Trägerplattform unter Wasser, können optische Verfahren auch zur Kartierung des Meeresbodens und natürlicher oder künstlicher Objekte in der Tiefsee eingesetzt werden, sofern die Messdistanz gering ist. Hydroakustische Methoden (Sound Navigation And Ranging, SONAR) sind die erste Wahl für mittlere Wassertiefen von 20–500 m (LURTON 2010). Während SONAR-Systeme auch für Vermessung großer Wassertiefen eingesetzt werden kann, bieten Satellitengravimetrie und -altimetrie globale Abdeckung von Tiefseegebieten (SANDWELL et al. 2014) mit einer räumlichen Auflösung im km-Bereich.

Im Flachwasserbereich besteht eine Überschneidung zwischen optischen Methoden und SONAR. SONAR übertrifft die optischen Methoden in Bezug auf die maximale Messtiefe. Die luftgestützte Erfassung mit optischen Methoden hat jedoch zwei Vorteile: (i) das effektive SONAR-Gesichtsfeld (engl. Field of View, FoV) verringert sich mit abnehmender Wassertiefe, während die Schwadbreite bei der Datenerfassung aus der Luft hauptsächlich von der Flughöhe und nur in geringem Maße von der Wassertiefe abhängt und (ii) für die schiffsgebundenen SONAR-Sensoren ist eine Mindestwassertiefe für den sicheren Betrieb erforderlich. Darüber hinaus bieten optische Methoden aus der Luft eine nahtlose Abdeckung vom Meeresboden über den Uferbereich bis zum trockenen ufernahen Bereich (GUENTHER et al. 2000; SCHWARZ et al. 2019; YANG et al. 2022).

Für hydrographische Vermessungen mittels optischer Fernerkundung haben sich folgende Methoden durchgesetzt: (i) radiometrische Tiefenbestimmung auf der Grundlage von Multispektralbildern (engl., Spectrally Derived Bathymetry, SDB), (ii) Mehrmedienphotogrammetrie auf der Grundlage von Stereobildern und (iii) luftgestützte Laserbathymetrie (engl. Airborne Laser Bathymetry, ALB). Während die ersten beiden Methoden passiv sind und die vom Boden des Gewässers zurückgestreute Sonnenstrahlung zur Tiefenmessung nutzen, ist ALB eine aktive Methode, die auf

Laufzeitmessung (engl. Time-of-Flight, ToF) eines gepulsten grünen Lasers basiert. zeigt eine schematische Darstellung der drei wichtigsten optischen Methoden in der Hydrographie.

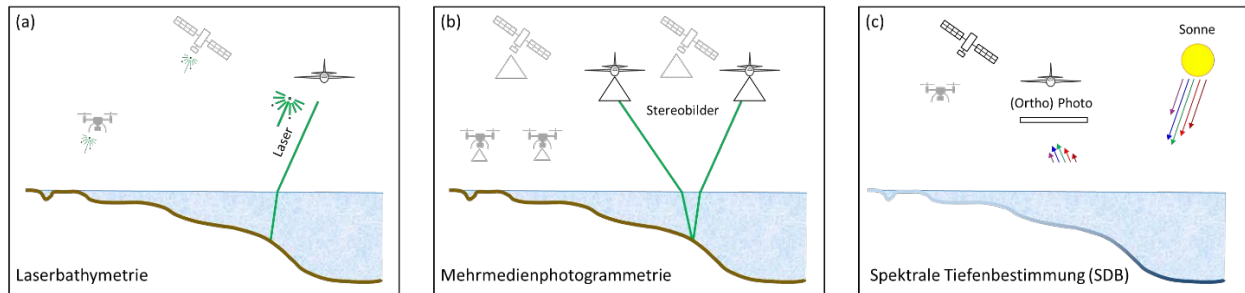


Abb. 2: Schematische Darstellung optischer Methoden in der Hydrographie: (a) Laserbathymetrie, (b) Mehrmedienphotogrammetrie, (c) spektrale Tiefenbestimmung. Alle drei Methoden können von Satelliten, bemannten und unbemannten Flugzeugen und auch unter Wasser eingesetzt werden.

Eine wegweisende Arbeit über Mehrmedienphotogrammetrie stammt aus dem Jahr 1948 (RINNER 1948). Die ersten Anwendungen des Lasers, der in den 1960er Jahren erfunden wurde, waren das Auffinden von Unterseebooten (SORENSEN et al. 1966) und die Kartierung der küstennahen Unterwassertopographie (HICKMAN & HOGG 1969). Nicht viel später beschrieben POICYN et al. (1970) multispektrale Ansätze zur Ableitung der Wassertiefe aus Satellitenbildern. Alle bekannten passiven und aktiven Methoden der optischen Hydrographie wurden also vor 1970 eingeführt, so dass die Wissenschaft auf Erkenntnisse aus 50 Jahren Forschung und Entwicklung aufbauen kann. Auch die Entwicklung von immer neuen Sensoren, Plattformen und Verarbeitungsstrategien treibt das Feld weiter voran. Dies zeigt sich deutlich an der steigenden Zahl von Veröffentlichungen in den letzten Jahren. Die Ergebnisse einer bibliografischen Abfrage in Scopus mit den Stichwörtern *spectrally derived bathymetry*, *multimedia photogrammetry* und *laser bathymetry* (einschließlich verwandter Begriffe) sind in Abb. 3 dargestellt.

Die kleine Spitze am Anfang der Zeitachse im Jahr 1982 ergibt sich aus der Zusammenfassung früherer Beiträge am Beginn der Grafik. Darüber hinaus ist aus Abb. 3a deutlich zu erkennen, dass bis etwa zum Jahr 2000 weniger als 20 Veröffentlichungen pro Jahr zu verzeichnen waren. Ab dem Jahr 2010 ist ein kontinuierlicher Anstieg zu verzeichnen. Interessant ist dabei, dass in der Anfangszeit (vor 1982) mehr Veröffentlichungen zum Thema Mehrmedienphotogrammetrie überwiegen, was höchst wahrscheinlich an der früheren Verfügbarkeit von photographischen (Stereo)-Kameras im Vergleich zu Multispektralsatelliten und bathymetrischen Laserscannern liegt. Für SDB ist anzumerken, dass Landsat Level-1-Multispektralbilder sowie Level-2- und Level-3-Wissenschaftsprodukte seit 2008 kostenlos aus dem Archiv des U.S. Geological Survey (USGS) heruntergeladen werden können, was ab 2010 zu einem kontinuierlichen Anstieg der SDB-bezogenen Publikationen führte (vgl. Abb. 3b). Bei der Laserbathymetrie ist aus Abb. 3d im Jahr 2011 eine Spitze zu erkennen, die auf das Aufkommen von topo-bathymetrischen Laserscannern zurückzuführen ist, die eine hohe räumliche Auflösung für Flachwassergebiete bieten.

Die auffälligen Spitzen in den Jahren 2014 und 2019 in den Abb. 3c und 3d können auf die zunehmende Verfügbarkeit von UAV-gestützten Bildsensoren zurückgeführt werden. Die stimulierende Wirkung von UAVs als Trägerplattformen ist ein allgemeiner Trend für photogrammetrische Kar-

tierung und 3D-Objektrekonstruktion (COLOMINA & MOLINA 2014). In Bezug auf hydrographische Anwendungen ist festzustellen, dass auch hier der Aufschwung bei der Mehrmedienphotogrammetrie früher einsetzte, da die Kameras leichter sind und es länger dauerte, kompakte und leichte Scanner zu entwickeln. Die hohe Zahl der Veröffentlichungen unterstreicht jedenfalls die Bedeutung der aktuellen Forschung, insbesondere vor dem Hintergrund des Klimawandels und seiner Auswirkungen, wie der Zunahme von Hochwasserkatastrophen einerseits und zunehmender Wasserknappheit und Dürren andererseits (KREIBICH et al. 2022).

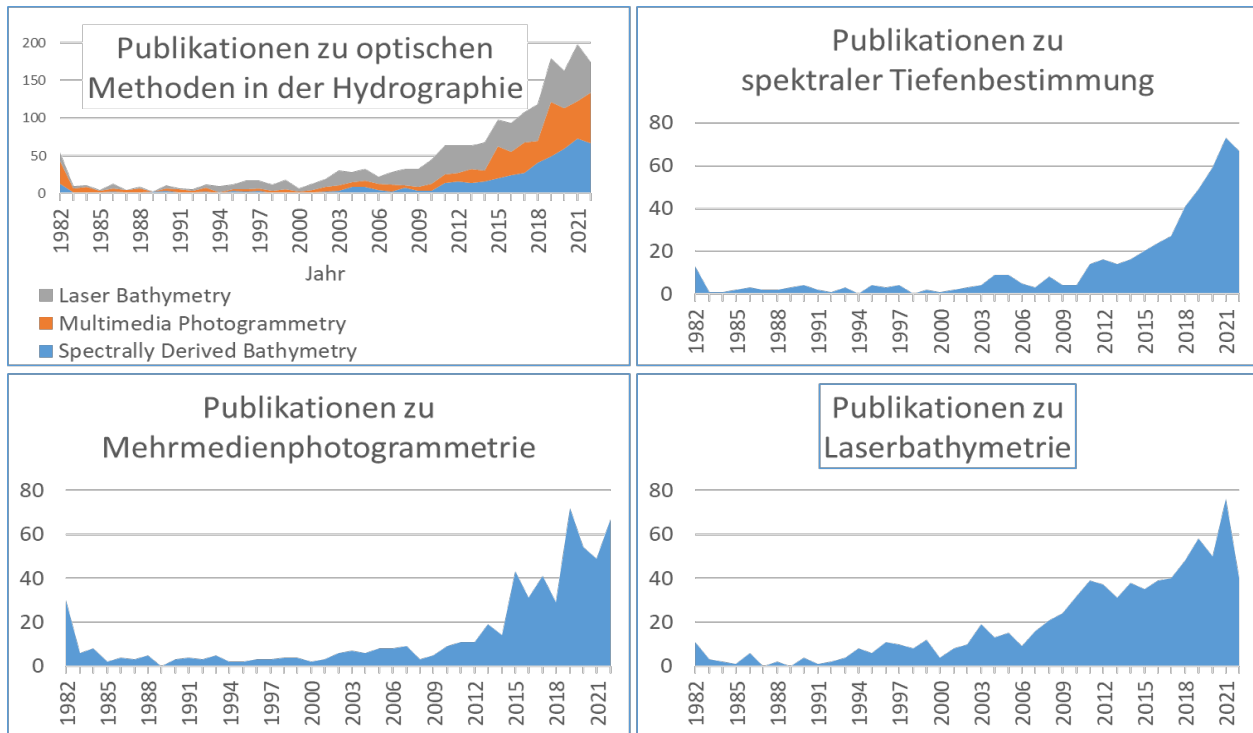


Abb. 3: Entwicklung der Veröffentlichungen im Zusammenhang mit optischen Methoden in der Hydrographie; Ergebnisse einer bibliographischen Abfrage in der Zitationsdatenbank Scopus.

Dieser Beitrag gibt einen Überblick über Methoden, Sensoren, Plattformen und Anwendungen optischer Messtechniken für die hydrographische Kartierung mit Schwerpunkt auf luftgestützter Erfassung. Der Artikel ist wie folgt aufgebaut: Im ersten Teil werden in Abschnitt 2 (SDB), Abschnitt 3 (Photobathymetrie) und Abschnitt 4 (Laserbathymetrie) die Prinzipien der drei wichtigsten optischen Methoden vorgestellt. Der zweite Teil des Artikels befasst sich in Abschnitt 5 mit Sensoren und Plattformen und stellt in Abschnitt 6 ausgewählte Anwendungen vor. Der Artikel endet mit einer Zusammenfassung und abschließenden Bemerkungen in Abschnitt 7.

## 2 Spektrale Tiefenbestimmung

Bei der spektral abgeleiteten Bathymetrie (SDB) wird eine Beziehung zwischen dem radiometrischen Bildinhalt und der Wassertiefe hergestellt (POICYN et al. 1970; LYZENGA 1978). Voraussetzung dafür ist ein umfassendes Verständnis der komplexen Wechselwirkung der Sonnenstrahlung mit der Atmosphäre, der Wasseroberfläche, dem Wasserkörper und schließlich dem Boden des

Wasserkörpers in Abhängigkeit von der Wellenlänge  $\lambda$ . Zur Ableitung der Bathymetrie aus dem radiometrischen Bildinhalt werden entweder physikalisch- oder regressionsbasierte Ansätze verwendet. Letzteres erfordert unabhängige Referenzdaten, um Modelle zu trainieren, welche die Beziehung zwischen Farbe und Tiefe beschreiben.

Grundlage ist in jedem Fall ein multispektrales Bild. Die gesamte am Bildsensor ankommende Strahlendichte kann als Summe der einzelnen Teilbeiträge geschrieben werden (LEGLEITER et al. 2009):

$$L_T(\lambda) = L_B(\lambda) + L_C(\lambda) + L_S(\lambda) + L_P(\lambda) \quad (1)$$

Die auf den Sensor einfallende Gesamtstrahlung  $L_T$  setzt sich aus der vom Gewässerboden reflektierten Strahlung  $L_B$ , der vom Gewässer bzw. der Wassersäule zurückgestreuten Strahlung  $L_C$ , der Signalkomponente aus Reflexionen an der Wasseroberfläche  $L_S$  und Komponenten aus rückstreuenden Partikeln in der Atmosphäre  $L_P$  zusammen. Die Signaldämpfung in der Wassersäule ist exponentiell, was auf die kontinuierliche Vorwärts- und Rückwärtsstreuung sowie auf die Absorption von Licht im Medium Wasser zurückzuführen ist. Der Signalbeitrag  $L_B$  vom Boden hängt sowohl von der Wassertiefe als auch von den Bodeneigenschaften (Reflexionsgrad, Rauheit) ab. Die spektralen Unterschiede des Bodens spielen allerdings nur in sehr flachem Wasser eine große Rolle, da in tieferem Wasser die Dämpfung durch die Wassersäule überwiegt (POPE & FRY 1997). Der Beitrag der Wassersäule  $L_C$  wird durch die optischen Eigenschaften des Wassers bestimmt. Wesentlicher Faktor dabei ist die Trübung durch Schwebstoffe und organische Stoffe (GROBELAAR 2009). Der Term  $L_S$  beschreibt die Signalantwort von der Wasseroberfläche. Ein großer Beitrag von der Oberfläche ist dann zu erwarten, wenn das Sonnenlicht durch spiegelnde Reflexion direkt in das Sichtfeld eines Sensorpixels einfällt.  $L_P$  schließlich repräsentiert die von der Atmosphäre in das Sichtfeld des Sensors zurückgestreute Strahlungsdichte.

Für Bilder, die optisch tiefes Wasser enthalten (kein messbares Bodensignal) kann eine einfache physikalische Beziehung zwischen Wassertiefe und Rückstreustärke formuliert werden (LYZENGA et al. 2006):

$$L(d) = L_S + L_B e^{-\alpha d} \quad (2)$$

$L(d)$  ist die am Sensor empfangene Strahlung nach Korrektur um die Atmosphäre und etwaige Glanzlichter. Der Term  $L_S$  umfasst sowohl die Reflexionen von der Wasseroberfläche als auch die Rückstreuung aus einer unendlich tiefen Wassersäule.  $L_B$  beschreibt in erster Linie das Reflexionsvermögen des Bodens, berücksichtigt aber auch Transmissionsverluste durch die Luft-Wasser-Grenzfläche und Effekte der Volumenstreuung im Wasserkörper. Der Exponentialkoeffizient  $\alpha$  ist der effektive Dämpfungskoeffizient und setzt sich aus der Summe von vorwärts und rückwärts gestreuten Lichtkomponenten zusammen. Gleichung 2 zeigt den exponentiellen Signalabfall in Abhängigkeit von der Wassertiefe  $d$  und dem Wassergehalt. Durch Entlogarithmieren kann eine lineare Beziehung zwischen der Wassertiefe und dem radiometrischen Bildinhalt hergestellt werden.

$$d = \frac{1}{\alpha} \ln \left( -\frac{L_B}{L(d) - L_S} \right) \quad (3)$$

Unter der Annahme, dass sowohl die Wasserverhältnisse als auch der Untergrund homogen sind, kann die Tiefe  $d$  also bereits aus einem einzigen spektralen Bildkanal bestimmt werden, ohne dass externe Referenzdaten vorhanden sind.

Der in Gleichung 3 formulierte Ansatz setzt homogene Reflektivität des Untergrunds voraus. Eine Methode, bei der Reflexionsunterschiede des Gewässerbettes teilweise kompensiert werden können, stammt von STUMPF et al. (2003). Dabei wird das Verhältnis von zwei Spektralbändern mit unterschiedlicher Wellenlänge berechnet, welches sich bei empirischen Untersuchungen als annähernd konstant erwies. Bei Multispektralbildern stehen in der Regel mehr als zwei Bänder zur Verfügung, sodass aus allen verfügbaren, wasserdurchdringenden Bändern die optimale Bandkombination ermittelt werden kann. Ein derartiges Verfahren wurde unter dem Namen Optimum Band Ratio Analysis (OBRA) von LEGLEITER et al. (2009) vorgestellt. Dabei ergibt sich eine aus dem Bild abgeleitete Größe  $X$ :

$$X = \ln \left[ \frac{L_T(\lambda_1)}{L_T(\lambda_2)} \right], \quad (4)$$

die annähernd linear mit der Wassertiefe  $d$  verbunden ist. Die empirische Beziehung zwischen  $X$  und  $d$  wird schließlich durch Regression von Referenztiefen aus Feldmessungen gegen  $X$  ermittelt. Ist der Gewässerboden komplex, reichen einfache Modelle basierend auf logarithmisch transformierten Bandenverhältnissen nicht mehr aus. Zusätzlich zu den oben erwähnten etablierten physikalischen und regressionsbasierten Tiefeninversionsmethoden haben daher in den vergangenen Jahren Ansätze des maschinellen Lernens (ML) für SDB zunehmend an Bedeutung gewonnen (AL NAJAR et al. 2021). Eine große Bedeutung kommt dabei speziell Convolutional Neural Networks (CNN) zugute, die sich aufgrund der Fähigkeit von Faltungen, beliebige Funktionen zu approximieren, als wirksame Werkzeuge für die spektrale Tiefenbestimmung erwiesen haben (ZHOU 2020).

### 3 Mehrmedienphotogrammetrie

Bei der Mehrmedienphotogrammetrie handelt es sich um eine rein geometrische Methode, deren Grundlagen auf die Mitte des 20. Jahrhunderts zurückgehen (RINNER 1948). Befindet sich der Sensor in der Luft oder am Boden und sind die zu vermessenden Objekte und Oberflächen untergetaucht, so ergibt sich ein Zwei-Medien-Problem. In der aktuellen Literatur wird die stereobildbasierte Erfassung der Unterwassertopografie für Fließgewässer (DIETRICH 2016) und für Küstengebiete diskutiert basierend auf Stereobildern, die mit Drohnen, bemannten Flugzeugen und Satelliten erfasst werden (HODÜL et al. 2018; CAO et al. 2019; AGRAFIOTIS et al. 2020).

Aufbauend auf dem Grundkonzept der Photogrammetrie (FÖRSTNER & WROBEL 2016) kann die Unterwassertopographie aus Stereobildern abgeleitet werden, sofern die innere und äußere Orientierung der Bilder bekannt ist (MULSOW 2010) und die Wasseroberfläche mit ausreichender Genauigkeit rekonstruiert werden kann. Sobald die innere und äußere Orientierung der Bilder aufgelöst ist und homologe Punkte des Wasserbodens in mindestens zwei Bildern identifiziert werden

$$\frac{\sin \alpha_a}{\sin \alpha_w} = \frac{n_w}{n_a} = \frac{v_a}{v_w}$$



können, muss der scheinbare Schnittpunkt der entsprechenden Bildstrahlen noch um die Brechung an der Wasseroberfläche korrigiert werden (LUHMANN et al. 2019). Die Grundlage hierfür ist das Snellius'sche Brechungsgesetz:

Gleichung 5 zeigt, dass der Sinus der luft- und wasserseitigen Bildstrahlen ( $\alpha_a$  und  $\alpha_w$ ) umgekehrt proportional zu den jeweiligen Brechungsindizes in Wasser ( $n_w \approx 1,33$ ) und Luft ( $n_a \approx 1,00$ ) und direkt proportional zu den Ausbreitungsgeschwindigkeiten sind ( $v_a \approx 300,000$  km/s,  $v_w \approx 225,564$  km/s). Abb. b veranschaulicht die Strahlbrechung. Die scheinbaren Bildpunkte, die sich aus dem geradlinigen Schnittpunkt der Bildstrahlen ergeben, sind tendenziell zu hoch und müssen durch Anwendung der Brechungskorrektur nach unten korrigiert werden. Ausführliche Beschreibungen der Refraktionskorrektur sind in KOTOWSKI (1988) und MURASE et al. (2008) zu finden. In Bezug auf die Bündelblockausgleichung beschreibt MAAS (2015) ein rigoroses geometrisches Korrekturmodell für den Multimediafall, das generisch genug ist, um als Grundlage für die Integration in Standard-Software für die photogrammetrische Bündelblockausgleichung zu dienen. In MAAS (2015) wird auch das Genauigkeitspotenzial der Mehrmedienphotogrammetrie erörtert und der Einfluss der Netzgeometrie, der Rauigkeit der Wasseroberfläche, der Schwankungen des Brechungsindex sowie von Dispersions- und Diffusionseffekten unter Wasser hervorgehoben, die unter relativ günstigen Bedingungen zu einer Verschlechterung der Genauigkeit um etwa einen Faktor zwei führen.

Für die Ableitung von 3D Punktwolken werden heutzutage Ansätze der merkmalsbasierten, flächenbasierten oder dichten Bildzuordnung eingesetzt. Im Bereich der dichten Bildzuordnung ist die Erzeugung von topographischen 3D-Punktwolken ausgereift. Offene Forschungsfragen gibt es dagegen noch in der Photobathymetrie (MANDLBURGER 2019). Dies gilt insbesondere für nicht statische Wasseroberflächen. Während bei topographischen Anwendungen eine Hinzunahme Stereobildpartnern i. d. R. die Genauigkeit der resultierenden Punktwolke erhöht, kann die richtungsabhängige Bildstrahlbrechung in der Photobathymetrie auch zu einer Verschlechterung der Ergebnisse führen. Das Problem der unterschiedlichen Refraktioneffekte für jede Punkt/Kamera-Kombination in einer Structure-from-Motion Punktwolke wird in DIETRICH (2016) behandelt. Der Autor stellt eine Multikamera-Refraktionskorrektur vor und berichtet von einer Genauigkeit in der Größenordnung von 0,1 % der Flughöhe, d. h. 4 cm bei einer Flughöhe von 40 m über Grund.

## 4 Laserbathymetrie

Im Gegensatz zu den in den Abschnitten 2 und 3 beschriebenen passiven Methoden handelt es sich bei der Laserbathymetrie um eine aktive Technik zur Kartierung flacher Gewässer unter Verwendung eines gepulsten grünen Lasers (PHILPOT 2019). Dabei wird die Entfernung zwischen Sensor und Ziel durch Messung der Umlaufzeit eines sehr kurzen Laserpulses (Wellenlänge  $\lambda = 532$  nm, Pulsdauer  $\Delta t = 1-10$  ns) durch Luft und Wasser bestimmt (GUENTHER et al. 2000). Nach dem Durchgang durch die Atmosphäre wird der ein Teil der Laserstrahlung an der Wasseroberfläche reflektiert, und der restliche Teil dringt in das Wasser ein. Beim Eintritt in die Wassersäule ändert der Laserstrahl in Abhängigkeit von den optischen Eigenschaften des Wassers sowohl seine Richtung als auch seine Ausbreitungsgeschwindigkeit gemäß dem Snellius'schen Brechungsgesetz (vgl. Gleichung 5,  $v_w \approx v_L/1,33 \approx 225,564$  km/s). Aufgrund der geringeren Lichtgeschwindigkeit

im Wasser erscheinen die unkorrigierten 3D-Unterwassermesspunkte im Gegensatz zur Mehrmedienphotogrammetrie zu tief und müssen entsprechend nach oben korrigiert werden.

In der Wassersäule wird die Laserstrahlung durch kontinuierliche Strahlbrechung und Signalabsorption abgeschwächt, so dass nach der Reflexion des Laserpulses am Boden und dem entsprechenden Rückweg nur noch ein kleiner Teil der Laserenergie am Sensor ankommt. Daher verwenden alle bathymetrischen Sensoren sehr empfindliche Detektoren (MANDLBURGER 2020). Die allgemeine Beziehung zwischen gesendeter und empfangener Energie wird durch die Laser-Radar-Gleichung (WAGNER 2010) beschrieben, die für bathymetrische Anwendungen in die Signalkomponenten der Wasseroberfläche, der Wassersäule, des Gewässerbodens und der Hintergrundstrahlung einschließlich der Verluste in der Atmosphäre unterteilt wird (ABDALLAH et al. 2012).

$$P_R = P_{WS} + P_{WC} + P_{WB} + P_{BK} \quad (6)$$

Gleichung 6 hat die gleiche Form wie Gleichung 1. Die Signalverluste in der Laserbathymetrie entsprechen somit denen, die bereits in Abschnitt 2 für die spektrale Tiefenbestimmung beschrieben wurden. Dies gilt auch für die exponentielle Abschwächung in der Wassersäule. Ein wesentlicher Vorteil der Laserbathymetrie besteht darin, dass die Signaldämpfung, die in der Regel durch den effektiven Dämpfungskoeffizienten  $k$  beschrieben wird, anhand der asymmetrischen Form der aufgezeichneten Wellenformen geschätzt werden kann (RICHTER et al. 2017; SCHWARZ et al. 2017). Es besteht eine direkte Beziehung zwischen  $k$  und der Secchi-Tiefe  $sd$  ( $sd \approx 1,6/k$ ). Die Secchi-Tiefe ist ein empirisches Maß für die Wassertrübung und bezeichnet den Abstand, ab dem die schwarzen und weißen Quadranten einer Scheibe von 20 cm Durchmesser, die von einem Boot an einem Seil ins Wasser gelassen wird, nicht mehr voneinander unterschieden werden können (EFFLER 1988). Die Hersteller von bathymetrischen Sensoren beschreiben die Leistung der Tiefenmessung in der Regel in Vielfachen der Secchi-Tiefe (MANDLBURGER 2020).

Die Erkennung der Wasseroberfläche ist eine Voraussetzung für eine präzise Brechung und Laufzeitkorrektur der Rohmessungen. Die meisten bathymetrischen Scanner betreiben zu diesem Zweck einen zusätzlichen Nahinfrarot-Kanal ( $\lambda = 1,064$  nm) zusammen mit dem grünen Laser, da die Signalabsorption im Wasser für Strahlung im nahen Infrarot (NIR) sehr hoch ist und NIR-Laserpulse daher nur minimal in die Wassersäule eindringen (GUENTHER et al. 2000). Wenn kein NIR-Kanal vorhanden ist, muss die Luft-Wasser-Grenzfläche allein anhand der Reflexionen des grünen Kanals modelliert werden. Da in diesem Fall die Echos von der Wasseroberfläche oft eine Mischung aus direkter Reflexion und Volumenstreuung in den ersten cm der Wassersäule darstellen (GUENTHER et al. 2000), sind spezielle Auswerte- und Modellierungsmethoden erforderlich. Insbesondere bei topo-bathymetrischen Scannern mit kleinem Laserfootprint müssen die Nicht-Planarität und die Dynamik der Wasseroberfläche (Wellen) berücksichtigt werden, um präzise 3D-Punktkoordinaten des Wasserbodens zu erhalten (WESTFELD et al. 2017).

Messunsicherheit, Ungenauigkeiten der Wasseroberfläche und des Refraktionskoeffizienten, tragen gemeinsam mit der Sensororientierung und -kalibrierung zum gesamten Fehlerbudget der Laserbathymetrie bei. Unter Anwendung der rigorosen Fehlerfortpflanzung veröffentlichten EREN et al. (2019) ein Modell zur Schätzung der vertikalen Gesamtunsicherheit (engl. Total Vertical Uncertainty, TVU) von bathymetrischem LiDAR.

Die meisten bathymetrischen Laserscanner arbeiten nach dem Prinzip der Laufzeitmessung. Vor allem für Unterwasseranwendungen im Nahbereich ist die Laserlichtschnitt-Triangulation eine Alternative zum ToF-Scanning. Bei der Lasertriangulation in der Hydrographie wird eine grüne Laserlinie auf das Objekt projiziert und die beleuchtete Linie bzw. Kurve wird von einer Kamera erfasst, die an einem festen Standort in Bezug zum Laserprojektor montiert ist (SARDEMANN et al. 2022). Das Abbildungssystem ist in einem wasserdichten Gehäuse installiert und die Sensoren müssen schräg im Gehäuse angebracht werden, um nahezu orthogonale Strahlenschnittwinkel zu erhalten. Um präzise 3D-Objektkoordinaten zu erhalten, müssen Brechungseffekte an der Luft-Glas-Grenzfläche im Inneren und an der Glas-Wasser-Grenzfläche außerhalb des Gehäuses berücksichtigt werden. Einer der Hauptvorteile gegenüber der ToF-basierten Laserbathymetrie ist, dass hochpräzise Unterwassermessungen zu wesentlich geringeren Kosten durchgeführt werden können. In der jüngeren Vergangenheit wurden verschiedene Implementierungen von bathymetrischen Scannern beschrieben, die auf dem Prinzip der Lichtblatt-Triangulation basieren (BLEIER et al. 2019), die alle eine Genauigkeit von weniger als einem Millimeter für einen begrenzten Tiefenbereich von weniger als 50 cm aufweisen.

## 5 Sensoren und Plattformen

Optische hydrographische Methoden werden sowohl auf globaler als auch auf lokaler Ebene eingesetzt. Der Messbereich variiert von 800 km bis zu wenigen Zentimetern. Die verwendeten Sensorplattformen werden entweder von weltraumgestützten, bemannten oder unbemannten luftgestützten, terrestrischen oder Unterwasserplattformen aus betrieben. Im Unterwasserbereich werden die Sensoren entweder von Tauchern, ferngesteuerten Fahrzeugen (ROV) oder autonomen Unterwasserfahrzeugen (AUV) getragen. Ein Überblick über die in der optischen Hydrographie verwendeten Plattformen und Skalen findet sich in CHEMISKY et al. (2021).

### 5.1 Weltraumgestützte Sensoren

Auf globaler Ebene sind multispektrale Satellitenbilder die wichtigste Quelle für die Erstellung hydrografischer Karten der Flachwasserzone. Diese werden sowohl für die weltraumgestützte Mehrmedienphotogrammetrie als auch für spektrale Tiefenbestimmung eingesetzt. Neben den Spektralkanälen bieten die meisten Sensoren auch einen panchromatischen Kanal mit höherer räumlicher Auflösung an. Die Pan-Kanäle sind für Photobathymetrie von Vorteil, liefern aber im Allgemeinen keine zusätzlichen Informationen für SDB. Die Spektralbänder umfassen in der Regel einen wasserdurchlässigen Coast Blue Kanal im ultravioletten Bereich des Spektrums ( $\lambda \approx 440$  nm), mehrere sichtbare Kanäle (blau, grün, rot, roter Rand) sowie Kanäle im nahen Infrarot (NIR), im kurzwelligen Infrarot (SWIR) und im thermischen Infrarot (TIR). Für die Ableitung hydrographischer Produkte bilden die NIR-Kanäle die Grundlage für die Korrektur des Sonnenlichts. Die sichtbaren Kanäle werden für die Ableitung der Bathymetrie verwendet (LYZENGA et al. 2006). Für die Mehrmedienphotogrammetrie werden Stereobilder benötigt, was die Auswahl an möglichen Satelliten einschränkt (z. B. WorldView, Pleiades, Terra/ASTER). Dennoch ist auch hier in den letzten Jahren ein gesteigertes Interesse zu beobachten, da diese rein geometrische Technik

voraussetzungsfrei ist und keine externen Referenzdaten benötigt. Anwendungen sind z. B. in HO-DÜL et al. (2018) beschrieben.

Während die Ableitung hydrographischer Produkte von weltraumgestützten Plattformen lange Zeit auf passive Bilder beschränkt war, hat sich diese Situation mit der Einführung von ICESat und seinem Nachfolger ICESat-2 (NEUMANN et al. 2019) grundlegend geändert. Der Satellit ist mit einem einzelphotonsensitiven Laserprofilier ausgestattet (Advanced Topographic Laser Altimeter System, ATLAS), der mit grüner Laserstrahlung ( $\lambda = 532 \text{ nm}$ ) arbeitet und damit neben seiner Hauptanwendung, der Erfassung der Kryosphäre der Erde, auch ideal für bathymetrische Zwecke geeignet ist. Obwohl ATLAS keine flächendeckende Flachwasserabdeckung mit hoher räumlicher Auflösung bietet, ergänzt der Sensor die bestehenden multispektralen Instrumente, indem er für jeden Laserpuls zuverlässige Unterwasser-Referenztopographie liefert. Dies ist besonders nützlich für Deep-Learning-basierte SDB-Ansätze, die eine große Menge an Trainingsdaten erfordern.

## 5.2 Luftgestützte Sensoren

Die klassische Anwendung der optischen Hydrographie erfolgt von bemannten Flugzeugen aus. Alle in den Abschnitten 2–4 besprochenen Erfassungsmethoden werden von bemannten Flugplattformen aus eingesetzt. Jede metrische Kamera, die üblicherweise für topographische Anwendungen und die Erstellung von Orthofotos verwendet wird, kann auch für die Ableitung hydrographischer Produkte verwendet werden. Dies gilt insbesondere für photogrammetrische Kameras, die neben den sichtbaren Kanälen auch einen NIR-Kanal bieten. Luftgestützte Laser-Bathymetrie-Sensoren können unterteilt werden in (i) tiefe bathymetrische Sensoren, (ii) flache topo-bathymetrische Sensoren und (iii) Mehrzweck-Sensoren. Bathymetrische Tiefensensoren zielen ab auf eine Maximierung der Eindringtiefe. Sie verwenden Laser mit einer relativ langen Pulsdauer von ca. 7 ns und einer niedrigen Messrate von 3–10 kHz, um eine hohe Pulsenergie von ca. 7 mJ zu erreichen. Um Sicherheitsvorschriften für Augensicherheit zu erfüllen, ist die Strahldivergenz solcher Sensoren groß (7 mrad), was bei einem Betrieb in 500 m Höhe zu einem Durchmesser der Laser-aufstandsfläche von 3–4 m führt. Die hohe Eindringtiefe von typischerweise 3 sd wird also mit einer mäßigen räumlichen Auflösung erkauft.

Bei den so genannten topo-bathymetrischen Sensoren liegt der Schwerpunkt auf einer höheren räumlichen Auflösung zur Erfassung von flachen Binnen- und Küstengewässern mit hoher Reliefenergie (Felsen, Geröll, plötzliche Hangänderungen usw.). Sie verwenden kurze und schmale Laserstrahlen (Pulsdauer: 1–2 ns, Strahldivergenz: 0,7–2 mrad) und höhere Pulswiederholraten von bis zu 700 kHz, was zu Laserabdruckdurchmessern von 0,5–1 m am Boden und einer Punktdichte von etwa 25 Punkten/m<sup>2</sup> in einem einzigen Flugstreifen führt. Die kurze Pulslänge ermöglicht die Trennung der Laserrückstrahlung von Wasseroberfläche und Boden auch bei sehr flachen Tiefzonen mit Wassertiefen unter 20 cm und damit einen nahtlosen Übergang zwischen Wasser und Land. Andererseits haben kurze Pulslängen auch eine geringere Pulsenergie und damit eine geringere Eindringtiefe von typischerweise 1,5 sd zur Folge.

Zusätzlich zu den Laserscannern enthalten die modernen Laserbathymetriescanner auch RGB- oder RGBI-Kameras. Die Bilder werden hauptsächlich zur Fotodokumentation oder als Datengrundlage für die Einfärbung von Punktwolken verwendet, aber der Einsatz von

hochauflösenden metrischen Kameras (z. B. RCD30, PhaseOne IXU usw.) eröffnet auch die Möglichkeit, neben der Laservermessung Bathymetrie auch mit Hilfe der Photogrammetrie sowie über spektrale Tiefenbestimmung abzuleiten. Für mehr Details sowie eine Liste verfügbarer Sensoren sei auf MANDLBURGER (2022) verwiesen.

### 5.3 UAV-gestützte Sensoren

Bis vor wenigen Jahren konnten bathymetrische Laserscanner aufgrund ihres hohen Gewichts nur von bemannten Plattformen (Flugzeuge, Hubschrauber, Tragschrauber) aus betrieben werden. Mit der fortschreitenden Miniaturisierung der Sensoren und Fortschritten bei der Entwicklung von unbemannten Flugplattformen können kompakte Laserscanner nun auch auf Starrflüglern und Multirotor-UAV integriert werden. Drohnen werden in der Regel aus einer niedrigen Flughöhe von etwa 50–120 m über dem Boden und mit einer moderaten Fluggeschwindigkeit von 4–10 m/s betrieben, was im Vergleich zum Betrieb von bemannten Flugplattformen in größerer Höhe eine deutlich kleinere Laserfläche sowie eine höhere Punktdichte und damit eine höhere räumliche Auflösung zur Folge hat. Außerdem ist die Signaldämpfung in der Atmosphäre aufgrund des kürzeren Messbereichs deutlich geringer und es steht effektiv mehr Signalstärke zur Verfügung, um in den Wasserkörper einzudringen. Dies gilt insbesondere für UAV-gestützte bathymetrische Lasersensoren, spielt aber auch bei der bildgestützten Bathymetrie eine Rolle.

Da leichte Kameras lange vor dem Aufkommen kompakter Laserscanner zur Verfügung standen, wurde der Einsatz von UAV-Kameras für photogrammetrische Kartierungen im Allgemeinen und hydrographische Anwendungen im Besonderen früher entwickelt als die UAV-gestützte Laserbathymetrie (DIETRICH, 2016). Wie bereits in Abschnitt 5.2 erwähnt, sind alle Kameras, die für die Kartierung der Topographie geeignet sind, auch für die Hydrographie geeignet. Während High-End-Kamerasysteme, die neben dem sichtbaren RGB-Kanal auch einen IR-Kanal enthalten, häufig für bemannte Plattformen in der Luft verfügbar sind, ist dies bei UAV-Bildern selten der Fall. Die aktuelle Forschung zeigt jedoch, dass RGB-Bilder eine geeignete Grundlage sowohl für die Medienphotogrammetrie als auch für SDB sind (siehe MANDLBURGER 2022).

In jüngster Zeit hat das Aufkommen von bathymetrischen Laserscannern, die in UAVs mit einer maximalen Startmasse (MTOM) von weniger als 35 kg integriert sind, einen weiteren großen Sprung im Bereich der luftgestützten Laserbathymetrie in Bezug auf die räumliche Auflösung und die Tiefenleistung bewirkt. So wie die Einführung von topo-bathymetrischen Scannern für flache Gewässer zusätzlich zu den traditionellen Tiefwassersensoren die räumliche Auflösung erhöht hat, haben UAV-basierte topo-bathymetrische Scanner die erreichbare Punktdichte um eine weitere Größenordnung gesteigert. Der Durchmesser der Laserfootprints moderner UAV-gestützter bathymetrischer Scanner liegt im sub-dm-Bereich, und zusammen mit Punktdichten in der Größenordnung von 100–200 Punkten/m<sup>2</sup> ermöglicht dies nicht nur die Kartierung der Unterwassertopographie mit hohem Detailierungsgrad, sondern auch die Erfassung und Modellierung von Unterwasservegetation und strömungsrelevante Mikrostrukturen wie z. B. kleinen Felsbrocken. Eine Liste verfügbarer Sensoren ist in MANDLBURGER (2022) enthalten.

## 5.4 Unterwasser-Sensoren

Obwohl sich der Artikel hauptsächlich auf optische hydrographische Methoden konzentriert, bei denen sich der Sensor oberhalb des Wasserspiegels befindet, seien hier auch kurz Unterwassersensoren und -plattformen diskutiert. Generell lassen sich vier verschiedene Szenarien unterscheiden, die in Abb. 4 schematisch dargestellt sind: (a) das Schiff oder Boot schwimmt auf der Wasseroberfläche und die Bildsensoren (Kameras und/oder Laserscanner) befinden sich auf der Unterseite des Schiffes in einem wasserdichten Gehäuse, (b) ein Taucher bedient manuell eine einzelne Kamera oder ein Stereokamera-Rig, (c) die Bildsensoren sind auf einem ferngesteuerten Fahrzeug (ROV) mit einer drahtgebundenen Kommunikationsverbindung integriert und (d) die Bildsensoren sind auf einem autonomen Unterwasserfahrzeug (AUV) installiert.

In jedem Fall ist der Abstand zwischen Sensor und Ziel relativ gering, was die Abbildung von Objekten in sehr hoher räumlicher Auflösung ermöglicht, aber zusätzliche Herausforderungen für die Sensororientierung mit sich bringt. Außer für den Fall des schwimmenden Schiffes befindet sich nicht nur der Sensor, sondern auch die Plattform vollständig unter Wasser, so dass GNSS für die Positionierung des Sensors nicht verfügbar ist. Wenn kein GNSS zur Verfügung steht, wird die Bildorientierung entweder über Passpunkte (MAAS 2015), visuelle Odometrie (BOTELHO et al. 2010) oder Simultaneous Localization and Mapping (SLAM)-Verfahren (BARKBY et al. 2009) erreicht. Der Einsatz von ROVs für Unterwasserinspektionen wird immer häufiger für die Kartierung und Überwachung von Offshore-Anlagen und Wasserkraftwerken eingesetzt. Neben ROVs entwickeln sich auch völlig autonom operierende Unterwasserfahrzeuge rasch. Sie werden bereits für die Kartierung großer Bereiche des Meeresbodens in Tiefen von mehreren tausend Metern eingesetzt. In einem ersten Überblick beschreibt BELLINGHAM (2009) die Funktionsweise und Navigation von AUV-Plattformen. Neben den hydroakustischen Sensoren sind in AUVs auch optische Bildsensoren wie Laser und Stereokameras integriert, aber die Anwendung wird unter anderem durch das Fehlen von Licht und durch Einschränkungen hinsichtlich des Energieverbrauchs behindert. Trotz dieser Schwierigkeiten wurden Kameras bereits erfolgreich für eine hochauflösende und großflächige Kartierung des Tiefseebodens eingesetzt (KWASNITSCHKA et al. 2016). Neben der Stereophotogrammetrie wird auch Laserscanning unter Wasser eingesetzt. Aus Gründen der Augensicherheit werden Scanner vorwiegend in ROVs und AUVs integriert. Unterwasser-Laserscanner arbeiten mit (i) dem ToF-Messprinzip auf der Basis von gepulsten grünen Lasern, (ii) Triangulation auf der Basis von strukturiertem Licht und (iii) Frequenzmodulation. Einen ausführlicheren Überblick zum Thema Unterwasser-Laserscanning geben FILISETTI et al. (2018) und MASSOT-CAMPOS & OLIVER-CODINA (2015).

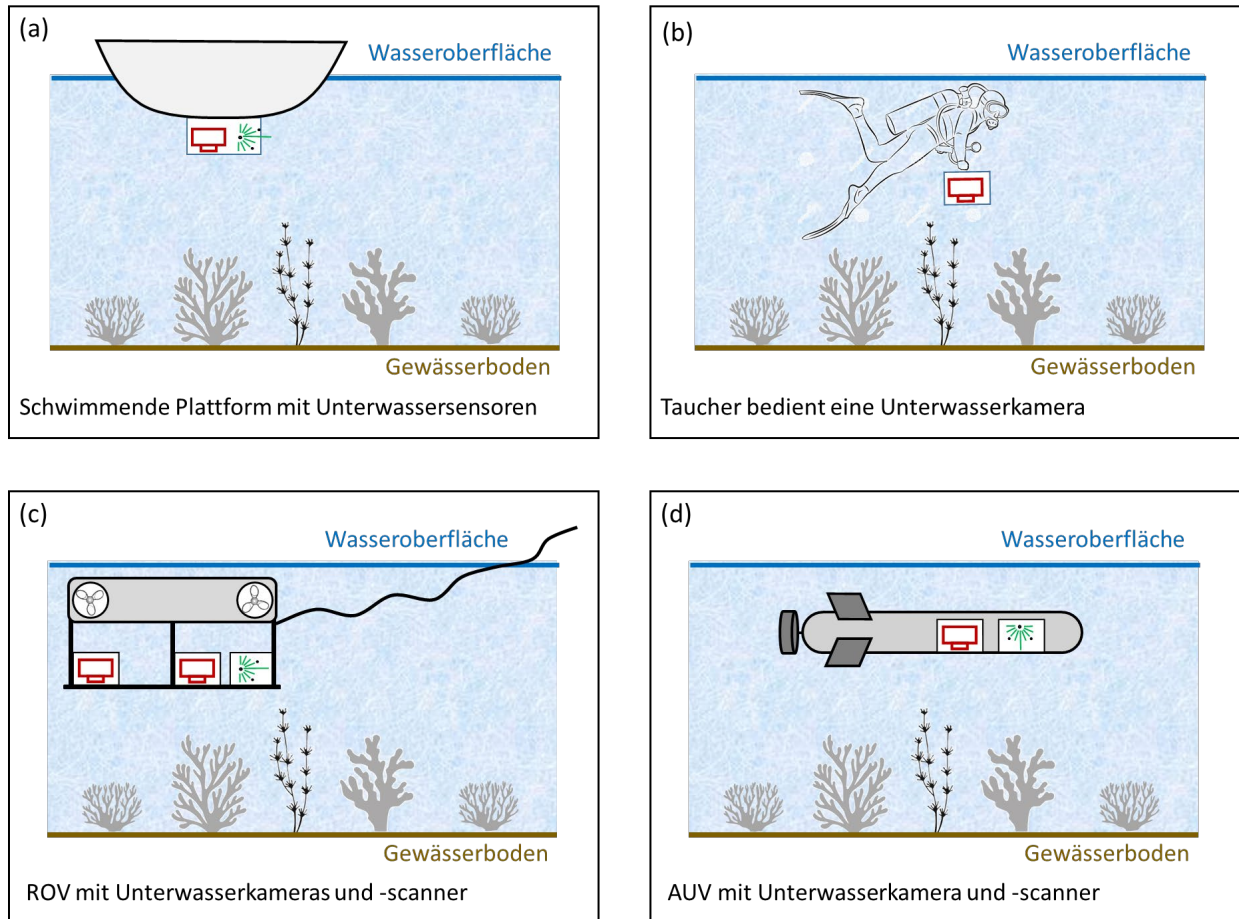


Abb. 4: Schematische Darstellung der Unterwassersensoren: (a) schwimmendes Schiff, (b) Taucher, (c) ferngesteuertes Unterwasserfahrzeug, (d) autonomes Unterwasserfahrzeug.

## 6 Anwendungen

Die Anwendungen von SDB, Photo- und Laserbathymetrie sind vielfältig, und aufgrund der enormen Fortschritte in der Sensor- und Plattformtechnologie entstehen laufend weitere Anwendungsszenarien. Insbesondere die Miniaturisierung der Sensoren und die Einführung von ferngesteuerten oder autonom operierenden Plattformen eröffnen neue Möglichkeiten für die Kartierung, Inspektion, Überwachung und Dokumentation von Unterwassertopographie, Artefakten und Infrastruktur. Es würde den Rahmen dieses Beitrags sprengen, auf alle Anwendungsbereiche einzugehen, im Folgenden werden daher lediglich beispielhafte Anwendungsfälle besprochen. Für eine ausführlichere Besprechung sei auch hier wieder auf [MANDLBURGER \(2022\)](#) verwiesen.

Optische Methoden eignen sich gut für die großflächige Kartierung von Flachwassergebieten mit moderaten Tiefen kleiner 60 m. Die effektivste Technik für großflächige Kartierungen ist die satellitengestützte spektrale Tiefenbestimmung. Ein globaler Ansatz wurde z. B. in [ALMAR et al. \(2021\)](#) veröffentlicht. Die Autoren behaupten darin, dass der Meeresboden bis zu einer Tiefe von 100 m erfasst und dabei die meisten Kontinentalschelfbereiche mit einer Fläche von 4.9 Mio. km<sup>2</sup>

kartiert werden konnten. Während die Tiefengenauigkeit von 6–9 m mäßig ist, ist die globale Abdeckung von besonderem Interesse für Länder, die nicht die Möglichkeit haben, In-situ-Messungen durchzuführen. Neben der SDB kann auch die luftgestützte Laserbathymetrie mit bemannten Flugzeugen eine großflächige Abdeckung mit einer viel höheren vertikalen Genauigkeit bieten, die den strengen IHO-Normen entspricht (IHO 2020). Die weite Verbreitung von bathymetrischem LiDAR wird durch die Verfügbarkeit von (offenen) Datenarchiven dokumentiert, die z. B. von der NOAA (National Oceanic and Atmospheric Administration) in den USA verwaltet und gepflegt werden (NOAA 2022).

Durch den Anstieg des Meeresspiegels sind viele archäologische Stätten über die Jahrhunderte überflutet worden. Dies gilt insbesondere für die römische Epoche, deren Spuren im Mittelmeer zu finden sind. Für die 3D-Rekonstruktion von Unterwasserstrukturen ist eine relativ hohe räumliche Auflösung erforderlich, weshalb topo-bathymetrisches LiDAR und multimediale Stereophotogrammetrie die bevorzugten Techniken sind. Wenn eine höhere räumliche Auflösung als die dm-Ebene erforderlich ist, ist die Unterwasserphotogrammetrie die Methode der Wahl (DRAP 2012). Interessant ist in diesem Kontext auch die Fusion von photogrammetrischen Datensätzen über und unter der Wasseroberfläche, die in NOCERINO & MENNA (2020) anhand eines prominenten Beispiels, dem Costa Concordia Schiffswrack, beschrieben wurde.

Da mehr als 200 Millionen Menschen an Küsten leben, die weniger als 5 m über dem Meeresspiegel liegen, besteht ein offensichtlicher Bedarf für die Kartierung von Küstengebieten mit dem Schwerpunkt auf dem Schutz dieser empfindlichen Übergangszone zwischen Meer und Land. Wenn globale Abdeckung und häufige Aktualisierungen wichtiger sind als eine hohe räumliche Auflösung, ist SDB die Methode der Wahl. Bei höheren Ansprüchen an Auflösung und Genauigkeit ist Laserbathymetrie weit verbreitet. Einige Länder haben bereits Erfassungsprogramme für den Küstenschutz und die Küstenüberwachung mit regelmäßigen Aktualisierungszyklen eingeführt, z. B. Schleswig-Holstein in Deutschland (CHRISTIANSEN 2021). In den U.S.A. wird die Überwachung von Küstenveränderungen seit Jahrzehnten mit dem Compact Hydrography Airborne Rapid Total Survey (CHARTS) Programm durchgeführt, bei dem bathymetrische und topographische Laserscanner sowie Luftbildkameras zum Einsatz kommen (MACON 2009).

Mit der Zunahme von Offshore-Installationen (Öl und Gas, Windturbinen usw.) wird die Inspektion, Kartierung und Überwachung von Unterwasser-Infrastrukturen immer wichtiger. Das gilt auch für Wasserkraftwerke. Um die zu Anlagen zu erfassen, werden sowohl ROVs als auch AUVs eingesetzt (CAPOCCI et al. 2017). Dabei wird Nahbereichsphotogrammetrie üblicherweise für die präzise Kartierung von Unterwasserinfrastruktur eingesetzt (CHEMISKY et al. 2021). Neben Stereokameras werden jedoch auch verschiedene Arten von Laserscannern eingesetzt (FILISSETTI et al. 2018; MASSOT-CAMPOS & OLIVER-CODINA 2015). Im Allgemeinen sind für die Inspektion und Überwachung von Infrastrukturen je nach Anwendung unterschiedliche Ansätze erforderlich, da manchmal eine sehr hohe Präzision im Sub-mm-Bereich erforderlich ist, um beispielsweise die Form einer Turbinenschaufel zu überprüfen, während es in anderen Fällen vielleicht nur notwendig ist, das Vorhandensein eines Hindernisses zu überprüfen (d. h. ein Bild oder eine Bildsequenz zu klassifizieren).



## 7 Zusammenfassung

In diesem Artikel wurde einen Überblick über optische Methoden in der Hydrographie gegeben und die eingesetzten Sensoren, Plattformen und typischen Anwendungen diskutiert. Die etablierten Methoden sind (i) spektrale Tiefenbestimmung, (ii) Mehrmedienphotogrammetrie und (iii) Laserbathymetrie. Alle drei Methoden können von weltraumgestützten, bemannten und unbemannten luftgestützten sowie Unterwasserplattformen aus betrieben werden.

Die spektrale Methode wird überwiegend auf der Grundlage von multispektralen Satellitenbildern verwendet, die den Vorteil haben, dass sie eine globale Abdeckung bieten. Für diese Methode sind externe Referenzdaten zur Kalibrierung der physikalischen oder regressionsbasierten Modelle erforderlich. Heute ersetzen Techniken des maschinellen Lernens zunehmend die traditionellen Methoden.

Das von Pilot:innen gesteuerte Flugzeug ist die bevorzugte Plattform für die Laserbathymetrie. Diese aktive Fernerkundungstechnik bietet eine gute Tiefenmessperformance von etwa der dreifachen Secchi-Tiefe, eine effiziente Flächenleistung mit dem Vorteil, dass die Schwadbreite nicht von der Wassertiefe, sondern nur von der Flughöhe abhängt, und einer sehr guten Positions- und Höhengenaugigkeit, die auch den strengen Spezifikationen der Internationalen Hydrographischen Organisation entspricht. Letzteres gilt insbesondere für die Flachwasserkanäle moderner topobathymetrischer Laserscanner, die eine hohe räumliche Auflösung im Submeterbereich und eine Tiefengenaugigkeit im dm-Bereich zum Preis einer reduzierten Tiefenleistung bieten. Dank der Fortschritte in der Sensor- und Plattformtechnologie können bathymetrische Laserscanner nun auf unbemannten Trägerplattformen integriert werden und bieten eine räumliche Auflösung und Genauigkeit im Submillimeterbereich.

Die Mehrmedienphotogrammetrie hingegen wird meist in der Unterwasservermessung eingesetzt, d. h. sowohl die Objekte als auch der Sensor befinden sich unter der Wasseroberfläche. Die Sensoren sind dabei in wasserdichten Gehäusen, die entweder flach oder kugelförmig sind. Von allen optischen Methoden hat die Mehrmedienphotogrammetrie die längste Geschichte, die auf eine wegweisende Arbeit aus 1948 zurückgeht. Heute wird die methodische Entwicklung von der Computer Vision vorangetrieben, was zur Einführung von Structure from Motion (SfM) und Dense Image Matching (DIM) in der Stereophotogrammetrie im Allgemeinen und der Mehrmedienphotogrammetrie im Besonderen geführt hat.

Zusammenfassend lässt sich sagen, dass optische Methoden eine effiziente Alternative zu herkömmlichen hydroakustischen Untersuchungen in flachen Gewässern darstellen. Beide Techniken ergänzen sich im Hinblick auf ihre jeweiligen Anwendungsbereiche. Für die Zukunft der optischen Methoden in der Hydrographie ist absehbar, dass kontinuierliche Fortschritte in der Sensor- und Plattformtechnologie einerseits und Fortschritte in den Verarbeitungsmethoden und der Computerleistung andererseits die Qualität der abgeleiteten Produkte weiter verbessern und auch neue Forschungsfelder erschließen werden. Insbesondere in Zeiten des Klimawandels werden multitemporale Analysen eine immer größere Rolle spielen. Dies ist für weltraumgestützte Daten mit entsprechenden Datenarchiven bereits gut etabliert, muss aber auf lokale hochauflösende Daten von luftgestützten, UAV-gestützten und Unterwasserplattformen ausgeweitet werden.

## 8 Literaturverzeichnis

- ABDALLAH, H., BAGHDADI, N., BAILLY, J.-S., PASTOL, Y. & FABRE, F., 2012: Wa-LiD: A New LiDAR Simulator for Waters. *IEEE Geoscience and Remote Sensing Letters Geosci. Remote Sensing Lett.*, **9**(4), 744-748, <http://dx.doi.org/10.1109/LGRS.2011.2180506>.
- AGRAFIOTIS, P., KARANTZALOS, K., GEORGOPOULOS, A. & SKARLATOS, D., 2020: Correcting Image Refraction: Towards Accurate Aerial Image-Based Bathymetry Mapping in Shallow Waters. *Remote Sensing*, **12**(2), <https://www.mdpi.com/2072-4292/12/2/322>.
- AL NAJAR, M., THOUMYRE, G., BERGSMAN, E. W. J., ALMAR, R., BENSILHA, R. & WILSON, D. G., 2021: Satellite derived bathymetry using deep learning. *Machine Learning*, <https://doi.org/10.1007/s10994-021-05977-w>.
- ALEVIZOS, E., OIKONOMOU, D., ARGYRIOU, A. V. & ALEXAKIS, D. D., 2022: Fusion of Drone-Based RGB and Multi-Spectral Imagery for Shallow Water Bathymetry Inversion. *Remote Sensing*, **14**(5), <https://www.mdpi.com/2072-4292/14/5/1127>.
- ALMAR, R., BERGSMAN, E. W. J., THOUMYRE, G., BABA, M. W., CESBRON, G., DALY, C., GARLAN, T. & LIFERMANN, A., 2021: Global Satellite-Based Coastal Bathymetry from Waves. *Remote Sensing*, **13**(22), <https://www.mdpi.com/2072-4292/13/22/4628>.
- BARKBY, S., WILLIAMS, S., PIZARRO, O. & JAKUBA, M., 2009: An efficient approach to bathymetric slam. 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, 219-224.
- BELLINGHAM, J., 2009: Platforms: Autonomous underwater vehicles. J. H. Steele (ed.), *Encyclopedia of Ocean Sciences* (2nd Edition). Academic Press, Oxford, 473- 484, <https://www.sciencedirect.com/science/article/pii/B978012374473900730X>.
- BLEIER, M., VAN DER LUCHT, J. & NÜCHTER, A., 2019: Scout3d – an underwater laser scanning system for mobile mapping. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, **42**(2/W18), 13-18, <https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLII-2-W18/13/2019/>.
- BOTELHO, S., OLIVEIRA, G., DREWS, P., FIGUEIREDO, M. & HAFFELE, C., 2010: Visual odometry and mapping for underwater autonomous vehicles. H. Yussuf (ed.), *Robot Localization and Map Building*. IntechOpen, Rijeka, Chapter 19, <https://doi.org/10.5772/9274>.
- CAO, B., FANG, Y., JIANG, Z., GAO, L. & HU, H., 2019: Shallow water bathymetry from WorldView-2 stereo imagery using two-media photogrammetry. *European Journal of Remote Sensing*, **52**(1), 506-521, <https://doi.org/10.1080/22797254.2019.1658542>.
- CHRISTIANSEN, L., 2021: Laser bathymetry for coastal protection in Schleswig-holstein. *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, **89**(2), 183-189, <https://doi.org/10.1007/s41064-021-00149-w>.
- COLOMINA, I. & MOLINA, P., 2014: Unmanned aerial systems for photogrammetry and remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, **92**, 79-97, <http://www.sciencedirect.com/science/article/pii/S0924271614000501>.
- DEWI, R. S., SOFIAN, I. & SUPRAJAKA, 2022: The application of satellite derived bathymetry for coastline mapping. *IOP Conference Series: Earth and Environmental Science*, **950**(1), 012088. <https://doi.org/10.1088/1755-1315/950/1/012088>.

- DIETRICH, J. T., 2016: Bathymetric Structure-from-Motion: extracting shallow stream bathymetry from multi-view stereo photogrammetry. *Earth Surface Processes and Landforms*, **42**(2), 355-364, <https://onlinelibrary.wiley.com/doi/abs/10.1002/esp.4060>.
- DRAP, P., 2012: Underwater photogrammetry for archaeology. D. C. da Silva (ed.), *Special Applications of Photogrammetry*. IntechOpen, Rijeka, Chapter 6, <https://doi.org/10.5772/33999>.
- EFFLER, S. W., 1988: Secchi disc transparency and turbidity. *Journal of Environmental Engineering* 114(6), 1436-1447. [https://doi.org/10.1061/\(ASCE\)0733-9372\(1988\)114:6\(1436\)](https://doi.org/10.1061/(ASCE)0733-9372(1988)114:6(1436)).
- EREN, F., JUNG, J., PARRISH, C. E., SARKOZI-FORFINSKI, N. & CALDER, B. R., 2019: Total Vertical Uncertainty (TVU) Modeling for Topo-Bathymetric LiDAR Systems. *Photogrammetric Engineering and Remote Sensing*, **85**(8), 585-596, <https://www.ingentaconnect.com/content/asprs/pers/2019/00000085/00000008/art00011>.
- FILISSETTI, A., MAROUCOS, A., MARTINI, A., MARTIN, T. & COLLINGS, S., 2018: Developments and applications of underwater LiDAR systems in support of marine science. *OCEANS 2018 MTS/IEEE Charleston*, 1-10.
- FÖRSTNER, W. & WROBEL, B. P., 2016: *Photogrammetric Computer Vision: Statistics, Geometry, Orientation and Reconstruction*, Springer International Publishing, Cham, Switzerland, 643-725, [http://dx.doi.org/10.1007/978-3-319-11550-4\\_15](http://dx.doi.org/10.1007/978-3-319-11550-4_15).
- GROBBELAAR, J., 2009: Turbidity. G. E. Likens (ed.), *Encyclopedia of Inland Waters*. Academic Press, Oxford, 699-704, <https://www.sciencedirect.com/science/article/pii/B9780123706263000752>.
- GUENTHER, G., CUNNINGHAM, A., LAROQUE, P. & REID, D., 2000: Meeting the accuracy challenge in airborne LiDAR bathymetry. *Proceedings of the 20th EARSeL Symposium: Workshop on LiDAR Remote Sensing of Land and Sea*. Dresden, Germany.
- HICKMAN, G. D. & HOGG, J. E., 1969: Application of an airborne pulsed laser for near shore bathymetric measurements. *Remote Sensing of Environment*, **1**(1), 47-58, <https://www.sciencedirect.com/science/article/abs/pii/S0034425769900881>.
- HODÜL, M., BIRD, S., KNUDBY, A. & CHÉNIER, R., 2018: Satellite derived photogrammetric bathymetry. *ISPRS Journal of Photogrammetry and Remote Sensing*, **142**, 268-277, <https://www.sciencedirect.com/science/article/pii/S0924271618301783>.
- IHO, 2020: S-44, Standards for Hydrographic Surveys, Edition 6.0, Standard 5th ed., International Hydrographic Organization, Monaco, [https://iho.int/uploads/user/pubs/standards/s-44/S-44\\_Edition\\_6.0.0\\_EN.pdf](https://iho.int/uploads/user/pubs/standards/s-44/S-44_Edition_6.0.0_EN.pdf), letzter Zugriff am 03.02.2023.
- IHO, 2022: Official website of the International Hydrographic Organization (IHO). <https://iho.int/en/>, letzter Zugriff am 03.02.2023.
- KOTOWSKI, R., 1988: Phototriangulation in Multi-Media-Photogrammetry. *Int. Arch. Photogramm. Remote Sens.* **27**, Kyoto, 324-334.
- KREIBICH, H., VAN LOON, A. F., SCHRÖTER, K. ET AL., 2022: The challenge of unprecedented floods and droughts in risk management. *Nature*, **608**(7921), 80-86, <https://doi.org/10.1038/s41586-022-04917-5>.
- KWASNITSCHKA, T., KÖSER, K., STICKLUS, J., ROTHENBECK, M., WEIB, T., WENZLAFF, E., SCHOENING, T., TRIEBE, L., STEINFÜHRER, A., DEVEY, C. & GREINERT, J., 2016: Deep-SurveyCam—A Deep Ocean Optical Mapping System. *Sensors*, **16**(2), <https://www.mdpi.com/1424-8220/16/2/164>.

- LEGLEITER, C. J., DAR, A. R. & RICK, L. L., 2009: Spectrally based remote sensing of river bathymetry. *Earth Surface Processes and Landforms*, **34**(8), 1039-1059, <https://onlinelibrary.wiley.com/doi/abs/10.1002/esp.1787>.
- LUHMANN, T., ROBSON, S., KYLE, S. & BOEHM, J., 2019: *Close-Range Photogrammetry and 3D Imaging*, De Gruyter, Berlin, Boston, <https://doi.org/10.1515/9783110607253>.
- LURTON, X., 2010: *An Introduction to Underwater Acoustics - Principles and Applications*, 2 ed., Springer-Verlag Berlin Heidelberg, Berlin, Heidelberg.
- LYZENGA, D. R., 1978: Passive remote sensing techniques for mapping water depth and bottom features. *Applied Optics*, **17**(3), 379-383.
- LYZENGA, D. R., Malinas, N. P. & Tanis, F. J., 2006: Multispectral bathymetry using a simple physically based algorithm. *IEEE Transactions on Geoscience and Remote Sensing*, **44**(8), 2251-2259.
- MAAS, H.-G., 2015: On the Accuracy Potential in Underwater/Multimedia Photogrammetry. *Sensors*, **15**(8), 18140-18152, <http://www.mdpi.com/1424-8220/15/8/18140>.
- MACON, C. L., 2009: USACE National Coastal Mapping Program and the next generation of data products. *OCEANS 2009*, 1-7.
- MANDLBURGER, G., 2019: Through-water dense image matching for shallow water bathymetry. *Photogrammetric Engineering and Remote Sensing*, **85**(6), 445-455.
- MANDLBURGER, G., 2020: A review of airborne laser bathymetry for mapping of inland and coastal waters. *Journal of Applied Hydrography*, **116**, 6-15.
- MANDLBURGER, G., 2022: A review of active and passive optical methods in hydrography. *International Hydrographic Review*, (28), 8-52, <https://doi.org/10.58440/ihr-28-a15>.
- MASSOT-CAMPOS, M. & OLIVER-CODINA, G., 2015: Optical Sensors and Methods for Underwater 3D Reconstruction. *Sensors*, **15**(12), 31525-31557. <https://www.mdpi.com/1424-8220/15/12/29864>.
- MULSOW, C., 2010: A flexible multi-media bundle approach. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, **36**(5), 472-477.
- MURASE, T., TANAKA, M., TANI, T., MIYASHITA, Y., OHKAWA, N., ISHIGURO, S., SUZUKI, Y., KAYANNE, H. & YAMANO, H., 2008: A photogrammetric correction procedure for light refraction effects at a two-medium boundary. *Photogrammetric Engineering and Remote Sensing*, **74**, 1129-1136.
- NEUMANN, T., MARTINO, A., MARKUS, T., BAE, S., BOCK, M. R., BRENNER, A. C., BRUNT, K. M., CAVANAUGH, J., FERNANDES, S. T., HANCOCK, D. W., HARBECK, K., LEE, J., KURTZ, N. T., LUERS, P. J., LUTHCKE, S. B., MAGRUDER, L., PENNINGTON, T. A., RAMOS-IZQUIERDO, L., REBOLD, T., SKOOG, J. & THOMAS, T. C., 2019: The Ice, Cloud, and Land Elevation Satellite – 2 mission: A global geolocated photon product derived from the Advanced Topographic Laser Altimeter System. *Remote Sensing of Environment*, **233**, 111325. <https://www.sciencedirect.com/science/article/pii/S003442571930344X>.
- NOAA, 2022: Coastal Topobathy Lidar. <https://coast.noaa.gov/digitalcoast/data/jalbtctx.html>, letzter Zugriff am 03.02.2023.
- NOCERINO, E. & MENNA, F., 2020: Photogrammetry: Linking the world across the water surface. *Journal of Marine Science and Engineering* **8**(2), <https://www.mdpi.com/2077-1312/8/2/128>.

- PHILPOT, W. (ed.), 2019: Airborne Laser Hydrography II, Cornell University Library (eCommons), Coernell. <https://ecommons.cornell.edu/handle/1813/66666>, letzter Zugriff am 03.02.2023.
- POICYN, F. C., BROWN, W. & SATTINGER, I. J., 1970: The measurement of water depth by remote sensing techniques, Technical Report 8973-26-F, Willow Run Laboratory, The University of Michigan.
- POPE, R. M. & FRY, E. S., 1997: Absorption spectrum (380–700 nm) of pure water. II. Integrating cavity measurements. *Appl. Opt.*, **36**(33), 8710-8723, <http://opg.optica.org/ao/abstract.cfm?URI=ao-36-33-8710>.
- RINNER, K., 1948: Abbildungsgesetz und Orientierungsaufgaben in der Zweimedienphotogrammetrie. *Österreichische Zeitschrift für Vermessungswesen*, Sonderheft (5).
- SANDWELL, D. T., MÜLLER, R. D., SMITH, W. H. F., GARCIA, E. & FRANCIS, R., 2014: New global marine gravity model from Cryosat-2 and Jason-1 reveals buried tectonic structure. *Science*, **346**(6205), 65-67, <https://science.sciencemag.org/content/346/6205/65>.
- SARDEMANN, H., MULSOW, C. & MAAS, H.-G., 2022: Accuracy Analysis of an Oblique Underwater Laser Light-sheet Triangulation System. *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, **90**(1), 3-18, <https://doi.org/10.1007/s41064-022-00196-x>.
- SCHWARZ, R., MANDLBURGER, G., PFENNIGBAUER, M. & PFEIFER, N., 2019: Design and evaluation of a full-wave surface and bottom-detection algorithm for LiDAR bathymetry of very shallow waters. *ISPRS Journal of Photogrammetry and Remote Sensing* **150**, <https://doi.org/10.1016/j.isprsjprs.2019.02.002>.
- SORENSEN, G., HONEY, R. & PAYNE, J., 1966: Analysis of the use of airborne laser radar for submarine detection and ranging, Technical report, SRI Report No. 5583, Stanford Research Institute.
- STUMPF, R. P., HOLDERIED, K. & SINCLAIR, M., 2003: Determination of water depth with high-resolution satellite imagery over variable bottom types. *Limnology and Oceanography*, **48**(1part2), 547-556, <https://aslopubs.onlinelibrary.wiley.com/doi/abs/10.4319/lo.2003.48.1part2.0547>.
- WAGNER, W., 2010: Radiometric calibration of small-footprint full-waveform airborne laser scanner measurements: Basic physical concepts. *ISPRS Journal of Photogrammetry and Remote Sensing*, **65**(6), 505-513, <http://dx.doi.org/10.1016/j.isprsjprs.2010.06.007>.
- WESTFELD, P., MAAS, H.-G., RICHTER, K. & WEIL, R., 2017: Analysis and correction of ocean wave pattern induced systematic coordinate errors in airborne LiDAR bathymetry. *ISPRS Journal of Photogrammetry and Remote Sensing*, **128**, 314-325, <http://dx.doi.org/10.1016/j.isprsjprs.2017.04.008>.
- YANG, F., QI, C., SU, D., DING, S., HE, Y. & MA, Y., 2022: An airborne LiDAR bathymetric waveform decomposition method in very shallow water: A case study around Yuanzhi island in the South China sea. *International Journal of Applied Earth Observation and Geoinformation*, **109**, 102788, <https://www.sciencedirect.com/science/article/pii/S0303243422001143>.

# AI-Based 3D-Detection of Parked Vehicles on a Mobile Mapping Platform using Edge Computing

JONAS MEYER<sup>1</sup>

*Zusammenfassung: Parkstatistiken bilden eine wichtige Grundlage für die Definition und Anpassung der Parkpolitik im städtischen Raum. Am Institut Geomatik der Fachhochschule Nordwestschweiz wurde das erste System entwickelt, welches basierend auf Low-Cost Sensoren die zuverlässige und kosteneffiziente Erstellung von Parkstatistiken für ganze Städte mit einer hohen Wiederholfrequenz erlaubt. Diese Lösung beruht jedoch vollständig auf zeit- und rechenintensivem Postprocessing und erfordert die Erfassung, Übertragung und Anonymisierung großer Datenmengen. In dieser Thesis wurde die bestehende Erfassungssoftware um die 3D-Fahrzeugdetektion erweitert. Dazu wurde unter Berücksichtigung der Low-Cost Komponenten eine passende 3D-Objektdetektionsmethode evaluiert und ein Workflow für die Datenvorverarbeitung entwickelt. Darüber hinaus wurde die 3D-Objektdetektionsmethode für die relevanten Fahrzeugklassen neu trainiert. Die weiterentwickelte Erfassungssoftware eliminiert sämtliche Einschränkungen bezüglich des Datenschutzes, reduziert das generierte Datenvolumen pro Bild um Faktor 60'000 von 15MB auf 0.25 KB und übertrifft mit einer Precision von 100% und einem Recall von 98% sämtliche bekannten Arbeiten für die Erstellung von Parkstatistiken auf der Basis von Mobile Mapping Daten.*

## 1 Einleitung

Wir erleben derzeit einen Wandel der städtischen Mobilität vom motorisierten Individualverkehr hin zu einer zunehmenden Vielfalt an multimodalen Mobilitätsangeboten. Durch Parkplätze belegter Straßenraum gerät aufgrund des ansteigenden Langsamverkehrs zunehmend unter Druck. Parkstatistiken bieten Behörden und politischen Entscheidungsträgern eine fundierte Grundlage für die Überprüfung und Anpassung der verfügbaren Parkplätze, der Parkregeln und -preise sowie der Parkpolitik im Allgemeinen. Die Erstellung von Parkstatistiken für Stadtteile oder ganze Städte ist jedoch nach wie vor ein größtenteils manueller Prozess. Die Parkstatistiken für die Stadt Basel wurden beispielsweise mithilfe von GoPro-Videos in Kombination mit der manuellen Interpretation durch menschliche Interpretation erstellt (RAPP TRANS AG BASEL-STADT 2019).

MATHUR et al. (2010), BOCK et al. (2015) und FETSCHER (2020) haben erfolgreich die Machbarkeit zuverlässiger Parkstatistiken mit Daten von Mobile Mapping Systemen (MMS) bewiesen. Allerdings sind diese Lösungen entweder abhängig von High-End MMS mit hochwertigen LiDAR und Kamera-Sensoren und daher teuer oder sie beschränken sich auf einen spezifischen Parkplatztyp. NEBIKER et al. (2021) entwickelten die erste Lösung zur zuverlässigen und kosteneffizienten Erstellung stadtweiter Parkstatistiken mit hoher Wiederholfrequenz (Abb. 1). Das System besteht aus einem elektrischen Dreirad, zwei RGB-D-Kameras *Intel RealSense D455* (INTEL CORPORATION 2020) (Abb. 1, a), einer GNSS/INS basierten Navigationseinheit (Abb. 1, b) und dem Em-

---

<sup>1</sup> FHNW Fachhochschule Nordwestschweiz, Institut Geomatik, Hofackerstrasse 30, CH-4132 Muttenz, E-Mail: jonas.meyer@fhnw.ch

bedded-System Modul *Nvidia Jetson TX2* (NVIDIA DEVELOPERS, 2022) (Abb. 1, c). Mithilfe künstlicher Intelligenz werden Fahrzeuge, in den aus Tiefenkarten gewonnenen Punktwolken, dreidimensional detektiert und daraus in einer GIS-Analyse Parkstatistiken abgeleitet. Diese Lösung beruht jedoch vollständig auf zeit- und rechenintensivem Postprocessing und erfordert die Erfassung, Übertragung und Anonymisierung großer Datenmengen. Um die Datenübertragung zu reduzieren, Einschränkungen aufgrund von Datenschutzbestimmungen zu beseitigen und geringe Latenzzeiten zu ermöglichen, werden Daten zunehmend in der Nähe der Erfassungsgeräte – an der Edge – verarbeitet. In den letzten Jahren wurden eigens dafür spezielle Hardware-Module mit einem geringen Energiebedarf entwickelt. Aufgrund der damit einhergehenden eingeschränkten Rechenressourcen, stellt die Kombination von Edge Computing und rechenintensiven Algorithmen wie die 3D-Objektdetektion (OD) gegenwärtig eine erhebliche Herausforderung dar.

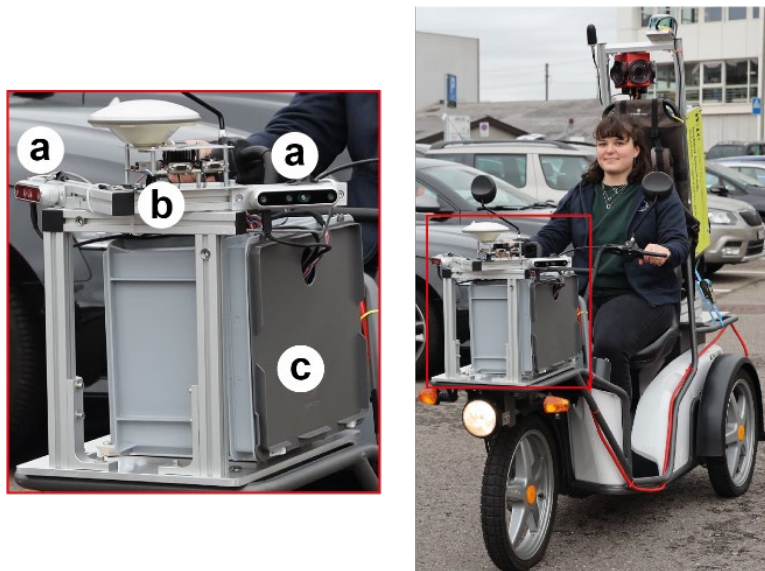


Abb. 1: Elektrisches Mobile Mapping System mit Low-Cost Sensoraufbau (Nebiker et al. 2021).

## 2 Materialien und Methodik

### 2.1 3D-Objektdetektor und Eingabedaten

Die Integration der 3D-Fahrzeugdetection direkt in das Erfassungssystem setzt nebst einer hohen Detektionsgenauigkeit eine tiefe Inferenzzeit voraus. NEBIKER et al. (2021) zeigten jedoch, dass moderne, auf dem KITTI Datensatz (GEIGER et al., 2012) trainierte 3D-OD-Methoden, angewendet auf die Punktwolken aus den RealSense (RS) Kameras, signifikant schlechtere Detektionsergebnisse und Inferenzzeiten erzielten als in den Originalpublikationen. Vergleiche beider Datensätze zeigten, dass die RS-Punktwolken ein massiv höheres Rauschen und teilweise große Datenlücken aufweisen sowie um Faktoren zwischen 200 und 400 dichter sind als die KITTI-Punktwolken (NEBIKER et al., 2021). Aus diesen Erkenntnissen wurden zwei Annahmen formuliert: 1) Eine Reduktion der Punktwolkendichte führt zu tieferen Inferenzzeiten und 2) durch Adaption der Charakteristiken der Trainingsdaten auf die RS-Punktwolken kann die Detektionsgenauigkeit gesteigert

werden. In zwei unabhängigen Experimenten wurden sieben Punktwolken-basierte 3D-OD-Methoden (Part-A<sup>2</sup> Net Anker-basiert und Anker-frei (SHI et al., 2020b), PointPillars (LANG et al. 2019), PointRCNN (SHI et al. 2019), PV-RCNN (SHI et al. 2020a), SECOND (YAN et al. 2018) und Voxel R-CNN (DENG et al. 2021)) hinsichtlich Detektionsgenauigkeit und Inferenzzeit evaluiert. Die Kombination der Resultate aus beiden Experimenten führte zur geeignetsten Methode sowie den optimalen Parametern für die Datenvorverarbeitung.

Für die Experimente mussten die RS-Punktwolken sowohl ausgedünnt als auch das Rauschen darin minimiert werden. Dazu wurde ein effizienter Datenvorverarbeitungsworkflow entwickelt, der anstelle von Punktwolken direkt die von den RS-Kameras erfassten Tiefenkarten verwendet (Abb. 2, links). Glättungsfilter können auf die als Bilddaten repräsentierten Tiefenkarten angewendet werden. Anschließend werden spezifische Pixel in der Tiefenkarte ausgewählt und via bekannte innere Orientierung der Kamera in 3D-Punkte umgerechnet. Ein regelmäßiges Gitter mit beliebiger horizontaler und vertikaler Schrittweite (Abb. 2, Mitte) selektiert die zu transformierenden Punkte. Dieser Ansatz erlaubt eine effiziente Generierung ausgedünnter und geglätteter Punktwolken (Abb. 2, rechts) bei gleichzeitiger Beibehaltung der Geometrieinformation und ohne aufwendige Szeneninterpretation.

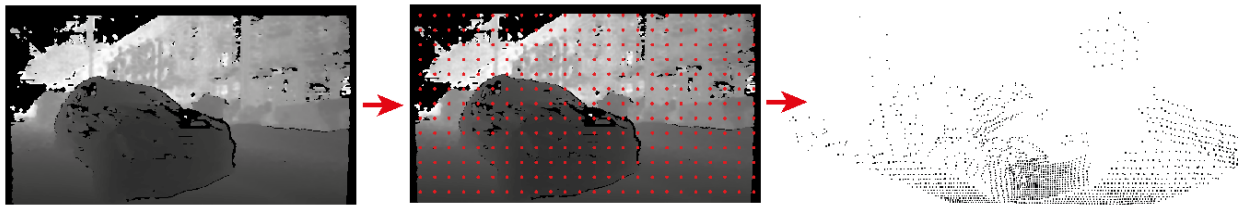


Abb. 2: Datenvorverarbeitungsworkflow von der Tiefenkarte zur aufbereiteten Punktwolke.

## 2.2 Training Objektdetektor

Für die Erstellung vollständiger Parkstatistiken wurde der 3D-Objektdetektor auf den Klassen „Car“, „Van“ und „Truck“ des KITTI Datensatzes (GEIGER et al. 2012) neu trainiert. Das neu trainierte Modell wurde auf den Testdaten des KITTI Datensatzes evaluiert und erreichte die in Tab. 1 dargestellten Average Precision (AP)-Werte bei einem IoU-Schwellwert von 0.7. Während die AP-Werte der Klasse „Car“ im selben Bereich wie jene des vortrainierten Modells liegen, deuten die AP-Werte der Klassen „Van“ und „Truck“ auf schlechte Trainingsresultate hin was aufgrund der stark unterrepräsentierten Trainingsdaten in diesen beiden Klassen zu erwarten war.

Tab. 1: Evaluationsresultate des neu trainierten Objektdetektors. AP<sub>70</sub>-Werte ausgewiesen in % pro trainierte Klasse und Schwierigkeitsstufe (einfach, moderat und schwierig).

Car			Van			Truck		
89.01	78.59	77.82	48.37	35.12	30.54	9.41	9.09	9.09

## 2.3 Erfassungssoftware

Die Erfassungssoftware von NEBIKER et al. (2021) beruht auf dem Graphen-basierten Robotik Framework Robot Operating System (ROS) (QUIGLEY et al. 2009) und verfügt über ein modulares Design. Die 3D-Fahrzeugdetektion konnte daher unkompliziert als neuer eigenständiger Knoten hinzugefügt werden (Abb. 3, rote Ellipse). Die weiterentwickelte Erfassungssoftware besteht aus



drei Knoten. Der *constrained-trigger* Knoten und der *realsense2-camera* Knoten (DORODNICOV & HIRSHBERG 2021) sind für das Triggering respektive die Publikation der Daten zuständig, während der *vehicle-detection* Knoten die 3D-Fahrzeugdetektion vollzieht (Abb. 3).

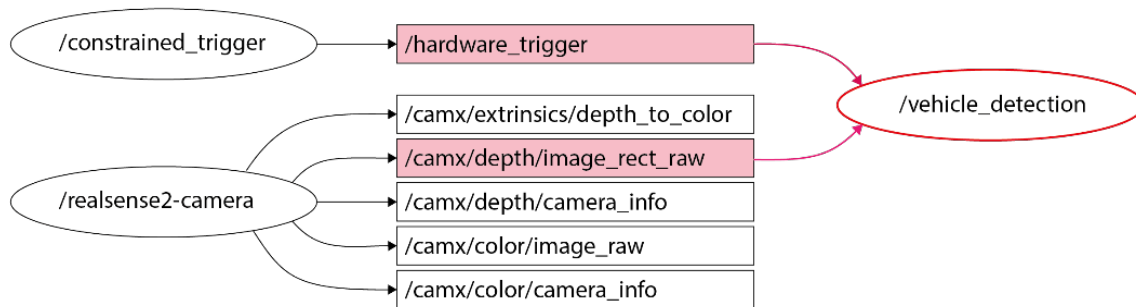


Abb. 3: ROS-Graph der erweiterten Erfassungssoftware. Knoten sind als Ellipsen und Topics als Vierecke dargestellt

Im *vehicle-detection* Knoten werden mit dem entwickelten Datenvorverarbeitungsworkflow und den evaluierten optimalen Parametern ausgedünnte Punktwolken aufbereitet, der neutrainierte Objektdetektor darauf angewendet und die Detektionsresultate gespeichert. Mit der Erweiterung der Erfassungssoftware entfällt die Notwendigkeit große Datenmengen zu speichern, zu übertragen und RGB-Bilder zu anonymisieren, wodurch sich die Anzahl der verbleibenden Nachbearbeitungsschritte auf die folgenden drei reduziert. 1) Berechnung der Bildposen mittels direkter Sensororientierung, 2) Transformation der Detektionsresultate via Bildposen ins übergeordnete Koordinatensystem, 3) GIS-Analyse zur Erstellung der Parkstatistik.

### 3 Experimente und Resultate

#### 3.1 3D-Objektdetektor und Eingabedaten

Mit dem eingeführten Datenvorverarbeitungsworkflow wurden insgesamt 83 Punktwolken-Datensätze mit unterschiedlich starken Glättungs- und Ausdünnungsfaktoren aufbereitet. Die Testdaten stammen aus einer Befahrung mit dem MMS von NEBIKER et al. (2021) und enthalten 593 RGB-D Bilder einer Quartierstrasse in der Stadt Basel mit 34 geparkten Autos. Sämtliche Datensätze wurden mit den sieben 3D-OD-Methoden auf einer Workstation ausgestattet mit einer performanten Grafikkarte prozessiert. Tab. 2 zeigt die erzielten Inferenzzeiten pro Methode und Ausdünnungsfaktor in Bilder pro Sekunden (FPS). Die Resultate zeigen deutlich, dass unabhängig von der 3D-OD-Methode spärlichere Punktwolken schneller prozessiert werden.

Tab. 2: Inferenzzeiten in FPS mit originalen und reduzierten Punktwolken (Reduktionsfaktor 196 & 400).

Methode	Part-A <sup>2</sup> Net Anker	Part-A <sup>2</sup> Net	PointPillars	PointRCNN	PV-RCNN	SECOND	VOXEL R-CNN
Original	7.6	3.6	22.3	3.5	1.7	15.5	7.2
196	11.4	13.0	44.5	4.6	10.9	28.9	20.9
400	11.6	13.2	45.1	4.9	11.1	29.0	22.3

Erste Detektionsresultate zeigten, dass eine vorgängige Glättung der Tiefenkarte keine signifikante Verbesserung der Detektionsresultate bewirkt. Außerdem benötigt die Glättung angesichts der geplanten Verarbeitung an der Edge zu viel Rechenleistung, weshalb die Reduktion des Rauschens in dieser Arbeit nicht weiterverfolgt wurde. Tab. 3 zeigt die besten Detektionsresultate pro 3D-OD-Methode gemessen am F1-Wert für die originale sowie die beste ausgedünnte Punktwolke zusammen mit dem entsprechenden Reduktionsfaktor.

Tab. 3: Beste Detektionsresultate auf ausgedünnten und originalen Punktwolken. Metrik F1-Wert

Methode	Part-A <sup>2</sup> Net Anker	Part-A <sup>2</sup> Net	PointPillars	PointRCNN	PV-RCNN	SECOND	VOXEL R-CNN
Original	0.2	0.3	-	<b>0.94</b>	0.56	0.46	0.34
Ausgedünnte PW (Red. Faktor)	<b>0.99</b> (180)	<b>0.99</b> (320)	0.76 (245)	<b>0.99</b> (500)	0.97 (320)	0.87 (245)	0.97 (180)

Basierend auf den erzielten Inferenzzeiten (Tab. 2) und Detektionsresultaten (Tab. 3) wurde der Objektdetektor Part-A<sup>2</sup> Net für den weiteren Verlauf dieser Arbeit verwendet.

### 3.2 Training Objektdetektor

Die in NEBIKER et al. (2021) publizierte Parkstatistik wurde mit dem neu trainierten Objektdetektor Part-A<sup>2</sup> Net wiederholt. Sie umfasst 350 Parkplätze, von denen 283 belegt waren. Die Fahrzeugdetektionen wurden analog zu NEBIKER et al. (2021) manuell ausgezählt und verifiziert. Tab. 4 zeigt die Ergebnisse aufgeschlüsselt nach Parkplatztyp. Der neu trainierte Objektdetektor erzielte über alle Parkplatz- und Fahrzeugtypen eine Precision von 100% und einen Recall von 98% und übertraf somit den von NEBIKER et al. (2021) verwendeten Objektdetektor um 11%. Es zeigte sich zudem, dass Fahrzeuge der Klassen „Van“ und „Truck“ zwar an der richtigen Stelle detektiert wurden, die geschätzten Dimensionen der 3D-Boundingboxen allerdings größtenteils falsch waren, was ebenfalls durch die schlechten AP-Werte in Tab. 31 angedeutet wird.

Tab. 3: Resultate der Parkstatistik unter der Verwendung des neu trainierten Objektdetektors. Resultate sind analog zu NEBIKER et al. (2021) nach Parkplatztyp aufgeschlüsselt.

	Parallel	Schräg	Senkrecht	2 x 2	Total
TP / TN / FP / FN	181 / 35 / 0 / 0	26 / 5 / 0 / 0	65 / 26 / 0 / 4	4 / 1 / 0 / 3	<b>276 / 67 / 0 / 7</b>
Precision	1.00	1.00	1.00	1.00	<b>1.00</b>
Recall	1.00	1.00	0.94	0.57	<b>0.98</b>

### 3.3 Erfassungssoftware

Die weiterentwickelte Erfassungssoftware wurde hinsichtlich Latenzzeit und generiertes Datenvolumen untersucht. Dazu wurde eine als ROS-Bag-Datei gespeicherte Kampagne erneut abgespielt und durch die Erfassungssoftware verarbeitet. Es konnte eine Latenzzeit von 0.53 s pro Bild (1.9 FPS) auf dem Embedded-System Modul Nvidia Jetson TX2 erreicht werden. Weiter wurden pro Aufnahme Daten von 0.25 KB generiert. Verglichen mit dem generierten Datenvolumen der alten Erfassungssoftware (ca. 15 MB / Aufnahme) wurde eine Reduktion um Faktor 60'000 erreicht.

## 4 Diskussion

In dieser Arbeit wurde die 3D-Fahrzeugdetektion in die in NEBIKER et al. (2021) beschriebene Erfassungssoftware integriert. Dazu wurden in verschiedenen Experimenten die einzelnen Komponenten evaluiert sowie das Gesamtsystem auf dessen Performance getestet. Die Experimente zeigten, dass die Eingabedaten einen erheblichen Einfluss sowohl auf die Detektionsgenauigkeit als auch die Inferenzzeit von 3D-OD-Methoden haben. Durch Ausdünnung der RS-Punktwolken konnten die Inferenzzeiten aller sieben evaluierten 3D-OD-Methode verringert werden (Tab. 2). Weiter konnten durch Adaption der Eigenschaften der Trainingsdaten auf die eigenen Daten deutlich bessere Detektionsresultate erzielt werden. Fünf von sieben 3D-OD-Methoden erreichten mit den ausgedünnten RS-Punktwolken einen F1-Wert von über 95%, währenddem nur eine von sieben Methoden auf den ursprünglichen Punktwolken einen F1-Wert von über 95% erreichte (Tab. 3). Hinsichtlich Detektionsgenauigkeit und Inferenzzeit erwies sich der Objektdetektor Part-A<sup>2</sup> Net als die beste 3D-OD-Methode für unseren Anwendungsfall.

Durch das Neutrainieren des Objektdetektors konnte eine sehr hohe Detektionsgenauigkeit (Precision 100%, Recall 98%) für alle Fahrzeugtypen erzielt werden erreicht (Tab. 4). Allerdings war die Schätzung der Fahrzeugdimensionen für die Klassen „Van“ und „Truck“ ungenügend. Als Hauptgründe werden hierfür die geringe Anzahl an Trainingsdaten sowie die sehr heterogenen Geometrien (Differenzen in Länge und Höhe von bis zu 2m) dieser beiden Klassen gesehen.

Die erweiterte Erfassungssoftware konnte die Einschränkungen des Vorgängersystem weitestgehend eliminieren und das generierte Datenvolumen um eindruckliche 60'000-Mal reduzieren. Allerdings zeigte die Performanceevaluation ein großes Verbesserungspotenzial bei der Latenzzeit, die im Durchschnitt bei 0.53 Sekunden (1.9 FPS) liegt. Die Latenzzeit ist stark von den verfügbaren Rechenressourcen des Nvidia Jetson TX2 abhängig und sollte daher mit einem neuen, leistungsfähigeren Hardwaremodul deutlich reduziert werden können. Darüber hinaus wäre es möglich, die Inferenzzeit des Objektdetektors durch Modellkomprimierung und Beschleunigungstechniken zu verringern, wie sie beispielsweise in NOUSIAS et al. (2021) demonstriert wurden.

## 5 Fazit und Ausblick

Durch die Detektion geparkter Fahrzeuge basierend auf den neutrainierten Objektdetektor Part-A<sup>2</sup> Net direkt auf dem Erfassungssystem und dem dazugehörigen Auswertungsworkflow konnten alle anderen bekannten Arbeiten zur Erstellung von Parkstatistiken (BOCK et al. 2015; FETSCHER 2020; GRASSI et al. 2017; MATHUR et al. 2010; NEBIKER et al. 2021) hinsichtlich der Detektionsgenauigkeit, der Anwendbarkeit auf verschiedene Parkplatztypen und einer hohen Wiederholffrequenz übertroffen werden. Darüber hinaus konnten die Einschränkungen des Vorgängersystem weitestgehend eliminiert werden. Für die effiziente Erstellung von Parkstatistiken muss die Erfassungssystem allerdings mittels neuerer Hardware noch weiter beschleunigt werden. Außerdem sollte die Detektion von Fahrzeugen, insbesondere die Schätzung der Dimensionen, verbessert werden, indem zusätzliche Trainingsdaten direkt mit unserem MMS gesammelt werden. Alternativ könnten auch andere öffentliche Trainingsdatensätze verwendet werden, was jedoch eine Anpassung der

evaluierten Parameter für die Datenvorverarbeitung erfordern würde. Schließlich würde eine robuste Echtzeit-Sensororientierung, die eine Position im Submeterbereich in herausfordernden städtischen Umgebungen sicherstellt, dass Postprocessing gänzlich überflüssig machen und eine Vielzahl zusätzlicher Anwendungsfälle ermöglichen.

## 6 Literatur

- BOCK, F., EGGERT, D. & SESTER, M., 2015: On-street Parking Statistics Using LiDAR Mobile Mapping. Proceedings of the 2015 IEEE 18th International Conference on Intelligent Transportation Systems, Gran Canaria, Spanien, 2812-2818. <https://doi.org/10.1109/ITSC.2015.452>.
- DENG, J., SHI, S., LI, P., ZHOU, W., ZHANG, Y & LI, H., 2021: Voxel R-CNN: Towards High Performance Voxel-based 3D Object Detection. Proceedings of the AAAI Conference on Artificial Intelligence, **35**(2), 1201-1209.
- DORODNICOV, S. & HIRSHBERG, D., 2021: realsense2\_camera. [http://wiki.ros.org/realsense2\\_camera](http://wiki.ros.org/realsense2_camera), letzter Zugriff 17.10.2022.
- FETSCHER, S., 2020: Automatische Analyse von Streetlevel-Bilddaten für das digitale Parkplatzmanagement. Bachelorarbeit, Fachhochschule Nordwestschweiz FHNW, Muttenz, Schweiz (unveröffentlicht).
- GEIGER, A., LENZ, P. & URTASUN, R., 2012: Are we ready for autonomous driving? The KITTI vision benchmark suite. Proceedings of the Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 3354-3361, <https://doi.org/10.1109/CVPR.2012.6248074>.
- GRASSI, G., JAMIESON, K., BAHL, P. & PAU, G., 2017: Parkmaster: An in-vehicle, edge-based video analytics service for detecting open parking spaces in urban environments. Proceedings of the Second ACM/IEEE Symposium on Edge Computing, San Jose, CA, USA, 1-14, <https://doi.org/10.1145/3132211.3134452>.
- INTEL CORPORATION, 2020: Intel®RealSense - Product Family D400 Series: Datasheet. <https://www.intelrealsense.com/wp-content/uploads/2020/06/Intel-RealSense-D400-Series-Datasheet-June-2020.pdf>, letzter Zugriff am 17.10.2022.
- LANG, A. H., VORA, S., CAESAR, H., ZHOU, L. & YANG, J., BEJBOM, O., 2019: Pointpillars: Fast encoders for object detection from point clouds. Proceedings of the Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 12689-12697, <https://doi.org/10.1109/CVPR.2019.01298>.
- MATHUR, S., JIN, T., KASTURIRANGAN, N., CHANDRASEKARAN, J., XUE, W., GRUTESER, M. & TRAPPE, W., 2010: ParkNet. Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services - MobiSys '10, San Francisco, CA, USA, 123, <https://doi.org/10.1145/1814433.1814448>.
- NEBIKER, S., MEYER, J., BLASER, S., AMMANN, M. & RHYNER, S., 2021: Outdoor Mobile Mapping and AI-Based 3D Object Detection with Low-Cost RGB-D Cameras: The Use Case of On-Street Parking Statistics. Remote Sensing, **13**(16), 3099, <https://doi.org/10.3390/rs13163099>.
- NOUSIAS, S., PIKOULIS, E.-V., MAVROKEFALIDIS, C., LALOS, A. S. & MOUSTAKAS, K., 2021: Accelerating 3D scene analysis for autonomous driving on embedded AI computing platforms.

- International Conference on Very Large Scale Integration (VLSI-SoC), Singapur, 1-6, <https://doi.org/10.1109/VLSI-SoC53125.2021.9606990>.
- NVIDIA DEVELOPERS, 2022: Jetson TX2 Module. <https://developer.nvidia.com/embedded/jetson-tx2>, letzter Zugriff 17.10.2022.
- QUIGLEY, M., CONLEY, K., GERKEY, B., FAUST, J., FOOTE, T., LEIBS, J., BERGER, E., WHEELER, R. & NG, A., 2009: ROS: an open-source Robot Operating System. ICRA, **3**(3.2) 5.
- RAPP TRANS AG BASEL-STADT, 2019: Erhebung Parkplatzauslastung Stadt Basel 2019. Basel, Schweiz.
- SHI, S., GUO, C., JIANG, L., WANG, Z., SHI, J., WANG, X. & LI, H., 2020a: PV-RCNN: Point-voxel feature set abstraction for 3D object detection. Proceedings of the Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 10526-10535, <https://doi.org/10.1109/CVPR42600.2020.01054>.
- SHI, S., WANG, X. & LI, H., 2019: PointRCNN: 3D object proposal generation and detection from point cloud. Proceedings of the Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 770-779, <https://doi.org/10.1109/CVPR.2019.00086>.
- SHI, S., WANG, Z., SHI, J., WANG, X. & LI, H., 2020b: From Points to Parts: 3D Object Detection from Point Cloud with Part-aware and Part-aggregation Network. IEEE Transactions on Pattern Analysis and Machine Intelligence, **1**, <https://doi.org/10.1109/tpami.2020.2977026>.
- YAN, Y., MAO, Y. & LI, B., 2018: Second: Sparsely embedded convolutional detection. Sensors, **18**(10), 3337, <https://doi.org/10.3390/s18103337>.

# Using Machine Learning to Predict Earth Deformation from InSAR Time Series

CHRISTOPH BAUMANN<sup>1</sup>, MOSTAFA KIANI SHAHVANDI<sup>1</sup>, ALEXANDRA DUCKSTEIN<sup>2</sup>,  
MALTE WESTERHAUS<sup>2</sup> & BENEDIKT SOJA<sup>1</sup>

*Abstract: Earth deformation can occur due to numerous causes and has the potential of far-reaching effects on society, infrastructure, and our natural surroundings. Forecasting its occurrence is crucial for deciding on preventative measures and increasing our understanding of the phenomenon. Using machine learning to predict earth deformation based on Interferometric Synthetic Aperture Radar time series, could be a promising approach for further research in order to facilitate these efforts. On a dataset focused on a former mining area, the selected approach denotes an average improvement of 41% for 12-day predictions compared to a baseline approach using quadratic regression. In addition to the recorded improvements, results suggest a higher noise resistance and greater generalization capabilities than the used baseline approach. While the results are promising, the derived model does likely not tap its full potential, leaving room for further research and improvement.*

## 1 Introduction

With the launches of satellites carrying radar technology, monitoring data of the earth's surface became available in near real-time with a high spatial resolution, allowing for millimeter scale monitoring even in the remotest of areas (ANANTRASIRICHAI et al. 2020). Pairing these characteristics with radar's inherent resistance to cloud coverage provides Interferometric Synthetic Aperture Radar (InSAR) with significant advantages in terms of costs and efficiency compared to other geodetic measuring systems. In the past, research efforts were largely focused on improving the processing methods of InSAR data (GONZÁLEZ et al. 2016). Thus, traditional time series (TS) analysis methods are still widespread and well-established methods to predict earth deformation. In this work, the capabilities of machine learning as an approach to predict earth deformation based on InSAR TS are to be investigated. Related work was performed by ANANTRASIRICHAI et al. 2020 and GRISHCHENKOVA 2018.

In the present setting, the main challenges lie in the length of the given TS for each individual point in space contained in the dataset and in the unknown noise level. The overarching goal of this work is therefore to determine the capabilities of an approach based on multilayer perceptrons (MLP) to predict earth deformation under these constraints. To investigate this, the metric of the mean absolute error (MAE) is used. In addition to that, a comparison to a baseline regression model is performed.

---

<sup>1</sup> ETH Zürich, Space Geodesy, Institute of Geodesy and Photogrammetry, Robert-Gnehm-Weg 15, CH-8093 Zurich, E-Mail: baumannch@ethz.ch

<sup>2</sup> Karlsruher Institut für Technologie, Geodetic Institute, Campus Süd, D-76128 Karlsruhe, E-mail: alexandra.duckstein@kit.edu

## 2 Data & Methodology

### 2.1 Data

The dataset at hand was captured between November 27<sup>th</sup>, 2015, and February 22<sup>nd</sup>, 2020 by the European Space Agency’s Sentinel 1A radar satellite on the descending orbit 139 and later processed at the Karlsruhe Institute of Technology. After the performed processing steps (MAZROOB SEMNANI et al. 2020), the dataset totals to 113 scenes. Each scene contains the line-of-sight (LOS) displacement for 218,637 persistent scatterers (PS) determined during prior processing steps. Not all 113 scenes are contained in the dataset, leaving unevenly spaced intervals. Having equal intervals in the time domain is an integral property for using the MLP-based approach. The missing scenes were therefore interpolated using linear interpolation.

The dataset contains an unknown level of noise. To quantify the course of each of the resulting 218,637 TS, a smoothness measure is introduced. This contrasts the TS to linear behavior by computing the absolute values of the second order derivatives, which then represents the respective degree of smoothness. A value of zero represents perfectly linear behavior. Small values thus imply smoother behavior in this context. As noted in table 3, a significant number of TS deviate from linear behavior to an extent that could suggest the presence of a certain noise level. Further conclusions however cannot be drawn.

### 2.2 Methods

The MLP was implemented in Python using the *Tensorflow* framework. The network consists of a series of dense layers. The number of neurons were guided by two fundamental principles. Either the layer sizes are multiples of 15 or powers of 2. In both cases, the layer constellations are to be symmetrical, excluding the input and output layers. In the model selection process, using a layer setup of 8-45-60-75-60-45-5 neurons in layers 1-7 prevailed and was selected for further investigations on the MLP’s capabilities. The detailed parameters of the model are summarized in table 1.

The model’s input length is set to 8 observations. This length is chosen due to the limitations implied by the length of the TS. Using an input length as mentioned above tries to account for the limited amount of information provided per location, while still attempting to maintain a significant level of information by containing a time interval of 96 days for each sample. The chosen output length of 5 facilitates 5 consecutive 12-day predictions. While this paper focuses on 12-day predictions, 60 days signifies the maximum prediction length capabilities of the selected model. In many cases, earth deformation occurs relatively slowly. Longer TS would therefore deliver more information on a location’s overall trend. The shortness of the individual TS has further implications. The model was trained on all TS simultaneously. This implies that the predictions made based on models of this kind are derived from the average behavior displayed in the dataset. The limiting factor of this type of approach is that points, where the trend deviates strongly from the average behavior, cannot be predicted accurately.

For testing purposes, 20% of all epochs were left aside for an independent assessment of the model’s predictions. Additionally, a validation split of 10% is applied. The hyperbolic tangent

function is used as activation function for the first 6 layers. For the last layer, a linear activation function is applied.

To facilitate the comparison to a baseline approach, a quadratic regression model is constructed. In order to match the MLP-based model’s input length, every TS is split into samples of the respective length. These samples are then fed individually in the quadratic regression model. The weights are thus computed for each individual sample ensuring a fair comparison between the two approaches.

Tab. 1: Layer setup and hyperparameter settings of the model

Layer Setup	Input Length	Output Length	Optimizer	Learning Rate	Batch Size
8-45-60-75-60-45-5	8	5	Adam	0.001	8,192

### 3 Results & Discussion

#### 3.1 Model Performance

Evaluating the MAEs for the entire dataset, a right-tailed behavior is observed with a mean value of 1.31mm and a standard deviation of 0.49mm. The respective indicators denote 1.49mm for the mean value and 0.77mm for the standard deviation, when only the independent test set is considered. Comparing the mean MAE to the median MAE of 1.32mm, a notable drop is apparent supporting the above-mentioned observation of a right-tailed distribution of MAEs. The median’s robustness to outliers suggests that there must be predictions where the MAE deviates significantly from what is observed for the majority of the points. Possible causes for such a phenomenon could be large prediction errors or high levels of noise on the respective parts of the data. The latter is more likely to be the root cause based on the presented findings. Ultimately, further analysis is required to assign a definite cause to the observed results.

Conclusively, it is crucial to note that while the MAE provides a good measure for assessing the quality of predictions, it does not respect commonly expected changes. In addition to that, the noise level for the dataset at hand is unknown, which limits the decisiveness of conclusions based on MAEs, since predictions only provide information about deformation trends if they exceed the noise level.

#### 3.2 Individual TS

The time series with indexes 51179 (TS51179) and 94809 (TS94809) show antagonistic degrees of smoothness. TS51179 is depicted in figure 1. The vertical red line represents the split between the training and the independent test set.

Tab. 2: Statistics on TS51179 and TS94809 for 1-step predictions on an independent test set

Index	MAE [mm]	Smoothness
51,179	6.67	12.42
94,809	0.06	0.05

TS51179 records a much higher MAE on the test set than TS94809 visualized in figure 2. The corresponding metrics are summarized in table 2. TS51179 oscillates rapidly, which correlates



with its smoothness indicator found in table 2, whereas TS94809, as suggested by its smoothness value, indeed follows a smoother pattern. The overall trend for both TS is observed to be non-linear.

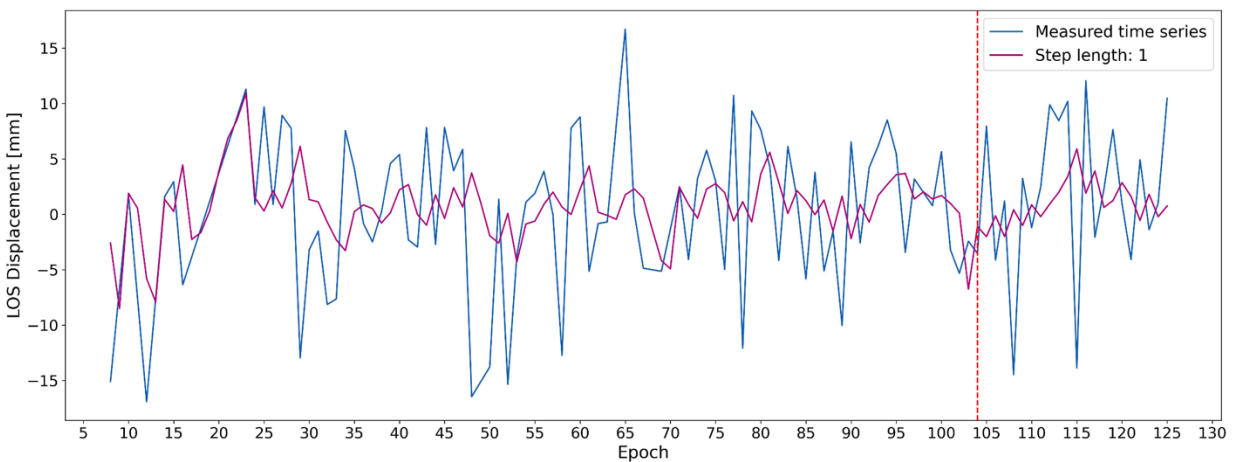


Fig. 1 TS51179: Measured vs. 12-day prediction (step length 1)

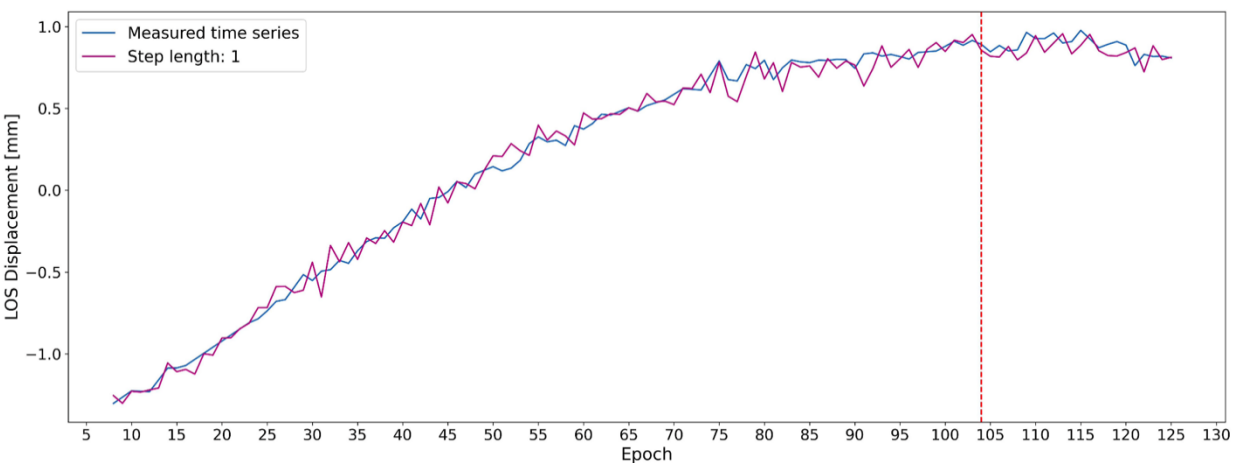


Fig. 2 TS94809: Measured vs. 12-day prediction (step length 1)

For TS51179, the 12-day prediction partly resembles a smoothed version of the measured TS. This phenomenon is most likely caused by the model's underlying assumption to treat all points indifferently. This results in the model being trained based on the average behavior, from which strongly fluctuating TS51179 deviates significantly at various points in time. Since TS51179 marks more rapid changes, the model is prone to underestimate those changes, resulting in a flattening of extreme values.

A similar causality is observed for the much smoother TS94809. Spikes are observed, e.g., at epoch 30. These spikes seem to follow increases or decreases in the input data to which the model overreacts and predicts exceeding amounts of change. This represents the counterpart to the above-described smoothing effect.

Another distinct artifact is the delayed prediction of minima and/or maxima. In the case of TS51179, this is clearly visible in figure 1 with a local maximum measured at approx. epoch 60. The respective local maximum is predicted by the model approx. 1 epoch later. Similar behavior is also observed around epoch 102 for a local minimum. In addition to that, this type of artifact is also found in the predicted TS of TS94809. They are most prominent for epochs 105 and 125. A prediction model's output can only change if there is indication in the data. This phenomenon occurs most distinctively for rapid changes. Delayed reactions to changes are relevant contributors to the MAE particularly for rapid and at the same time large changes, since there the predicted deformation deviates the most. These delays in predicting local extrema are typical forecasting artifacts. This effect can therefore be decoupled from the MLP-based approach.

### 3.3 Comparison to Baseline Approach

The quadratic regression model scores an average MAE of 2.23mm for 12-day predictions. This corresponds to a 70.81% increase compared to the in this work developed model's average MAE of 1.31mm. Table 3 summarizes the achieved MAE per smoothness category for both the neural network and quadratic regression.

Even though the MLP-based approach shows better performance for all bins, the lowest improvements are denoted for the smoothest bins. These TS most resemble linear or quadratic behavior. For all other bins, the MLP outperforms quadratic regression by a significant margin. These two facts combined suggest that the MLP-based approach could be more noise resistant than the baseline approach using quadratic regression. This suggested property could be invaluable since even after the most sophisticated processing steps, there will most certainly still be noise present. Though, this does not imply sufficient prediction accuracies, it could suggest a higher robustness to noise. Besides the fact that the MLP-based model achieves significantly better MAEs on average, it also offers greater generalization capabilities. Also, while a quadratic regression model needs to re-estimate its parameters for each input, resulting in immense computational costs, the MLP, once it has been trained, is fast to be applied even for large data sets.

Tab. 3: Comparison of quadratic regression (QR) and the selected MLP-based approach in regard to prediction accuracies for 12-day predictions

<b>Lower Bound</b>	<b>Upper Bound</b>	<b>Number of TS per bin</b>	<b>MAE QR [mm]</b>	<b>MAE MLP [mm]</b>	<b>Improvement [%]</b>
<i>None</i>	0.5	427	0.14	0.13	8.16
0.5	1.0	5,436	0.76	0.47	38.03
1.0	2.0	62,903	1.37	0.82	40.12
2.0	3.0	78,331	2.14	1.26	41.30
3.0	4.0	48,575	2.93	1.70	42.03
4.0	5.0	17,938	3.67	2.12	42.29
5.0	<i>None</i>	5,027	4.43	2.56	42.38

## 4 Conclusions & Outlook

Though no absolute prediction accuracy can be determined in this study, based on the investigated metrics, the MLP-based model concludes to be a promising approach for future research. Achieving an average improvement of 41% for 12-day predictions in comparison to the quadratic-regression-based approach suggests that the MLP-based approach provides greater generalization capabilities.

To further assess the MLP-based approach's generalization capabilities, comparisons to other statistical prediction techniques would be beneficial. The developed model is only one of a vast number of possible layer constellations. In future work, further setups could be investigated, perhaps leading to more accurate results. Based on the recorded success of the deep learning technique used in this study, investigating more sophisticated deep learning approaches, such as long short-term memory networks, may also provide opportunities for advances in the field.

## 5 References

- ANANTRASIRICHAI, N., BIGGS, J., KELEVITZ, K., SADEGHI, Z., WRIGHT, T., THOMPSON, J., ACHIM, A. & BULL, D., 2020: Deep learning framework for detecting ground deformation in the built environment using satellite InSAR data., <https://doi.org/10.48550/arXiv.2005.03221>.
- GONZÁLES, P., WALTERS, R., HATTON, E., SPAANS, K., HOOPER, A. & WRIGHT, T., 2016: LiCSAR: An Automatic InSAR Tool for Measuring and Monitoring Tectonic and Volcanic Activity. *Remote Sensing*, **12**(15), 2430, <https://doi.org/10.3390/rs12152430>.
- GRISHCHENKOVA, E.N., 2018: Development of a neural network for earth surface deformation prediction. *Geotech. Geol. Eng.*, **36**, 1953-1957, <https://doi.org/10.1007/s10706-017-0438-y>.
- MAZROOB SEMNANI, N., BREUNIG, M., AL-DOORI, M., HECK, A., KUPER, P. & KUTTERER, H., 2020: Towards intelligent geo-database support for earth system observation: improving the preparation and analysis of big spatio-temporal raster data. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **43**(B4-2020), 485-492, <https://doi.org/10.5194/isprs-archives-XLIII-B4-2020-485-2020>.

# Evaluation of Historical Aerial Images in a Landslide Area

ADRIAN WENG<sup>1</sup>, FRANZISKA WILD<sup>2</sup> & EBERHARD GÜLCH<sup>3</sup>

*Abstract: Cadastral photogrammetry was approved in Baden-Württemberg in 1971 and was used in the 1970s and 1980s to determine coordinates for the real estate cadastre from land consolidation areas. In these former land consolidation areas, the boundary and survey points of the real estate cadastre today have an inferior coordinate quality. In the Eberstadt I R procedure from 1972 a photogrammetric re-evaluation has been carried out 50 years after Prof. Fritz Ackermanns early innovations in cadastral photogrammetry with the aim of determining final UTM coordinates.*

*A workflow for photogrammetric re-evaluation of historical aerial images was being developed by the surveyor's office Heilbronn and applied in the Eberstadt I R project. In addition, the known landslide has been investigated. This involves comparing a current terrain model to a historical terrain model, which results in elevation differences for the landslide area. In the Eberstadt I R procedure, it was thus possible to locate a clear break-off edge of the landslide and to precisely delimit an 87,500 m<sup>2</sup> landslide area with changes in elevation.*

## 1 Introduction

Since 2021 the Surveying Office of the county of Heilbronn develops methods to evaluate historical aerial images which were used for land consolidation in the 70s and 80s of the last century and were also part of Prof. Fritz Ackermanns early innovations in cadastral photogrammetry. Cadastral photogrammetry has been applied to scope with the huge number of new real estate cadastre- and surveying points in the land consolidation projects. In the state of Baden-Württemberg cadastral photogrammetry has been applied between 1970 and 1983 in 142 project areas. In the county of Heilbronn approximately 60 historical projects for vineyard clearing are known. About 6.000 to 7.000 surveying points have been determined as well as between 80.000 and 100.000 real estate boundary points. These points have been determined in the former Soldner coordinates of the former states of Württemberg and Baden. Due to the geodetic network stresses they are only available in the lower coordinate quality level UTM-G (graphical coordinates) as described by (WILD-PFEIFFER 2018).

The aim of the recent developments at the Heilbronn Surveying Office is to generate cadastre-compliant state coordinates by re-evaluating the historical aerial images. This requires not only the aerial images and historical camera and flight data, but also control points signalized at the time, whose markings are still unchanged today and can thus be determined on site by surveying.

The project area Eberstadt I R has been selected, as in the western part of the area landslides are suspected. These landslides are known from past surveys of the Heilbronn survey office in this

---

<sup>1</sup> Siemensstraße 240, D-53121 Bonn, Germany, E-Mail: adrian.weng@web.de

<sup>2</sup> Vermessungsamt am Landratsamt Heilbronn, Lerchenstraße 40, D-74072 Heilbronn, E-Mail: franziska.wild@landratsamt-heilbronn.de

<sup>3</sup> Stuttgart University of Applied Sciences (HFT Stuttgart), Center for Geodesy and Geoinformatics, Schellingstraße 24, D-70174 Stuttgart, Germany, E-Mail: eberhard.guelch@hft-stuttgart.de

area, since extreme network stresses were repeatedly noticed. Some of the landslides amount to more than 2 meters downslope and can even be recognized in current orthophotos. However, the extent and exact area of the landslides is not known and was supposed to be clearly defined by the bachelor thesis of the principal author.

For a new evaluation, the analogue aerial photos must be scanned with an aerial photogrammetric scanner and be available digitally. In addition, a sufficient number of control points must be available, which are signaled in the aerial images and determined by surveying in the field. Additional data, such as camera calibration, has to be collected for the initial situation. The re-evaluation of the aerial images is performed with Erdas Imagine. After the interior orientation, an aerial triangulation is evaluated to subsequently calculate coordinates for signaled points. A specified accuracy of the coordinates has to be achieved that they are suitable for transfer to the real estate cadastre. For the detection of the landslide area further photogrammetric methods and software modules for the processing of point clouds are used.

## **2 Development of Cadastral Photogrammetry in the State of Baden-Württemberg**

In order to understand the background of cadastral photogrammetry, the chronological classification is especially important. The development of cadastral photogrammetry in Baden-Württemberg is shown in Figure 1.

**Before 1960:** Until 1920, the application of photogrammetry was mostly focused on graphical results, with exceptions. In the period between 1925 and 1960, instrumental developments became the main focus. Photogrammetric tasks such as aerial triangulation or point determination were solved graphically or mechanically where possible (ACKERMANN 1973).

**Between 1960 and 1970:** The shift to numerical photogrammetry began around 1960. Causes for the shift were particularly strong developments in the area of computer systems, data processing and the corresponding equipment developments for the measurement of model coordinates. In addition, new tasks arose for photogrammetry. For example, cadastral photogrammetry in Baden-Württemberg came into focus, which demanded high accuracies. Nevertheless, the introduction of numerical methods stagnated for a decade. The transition to the technical possibilities progressed only slowly and the factors of hardware and software components were misjudged.

However, aerial triangulation and photogrammetric orientation were extensively tested during this period. Improved hardware (cameras, films, measuring instruments) and programs for computational block adjustment allowed this development to gain momentum (ACKERMANN 1973).

**Between 1970 and 1973:** At this point, mainly the development of cadastral photogrammetry in Baden-Württemberg is in focus, which was primarily driven by the Institute for Photogrammetry at the University of Stuttgart. In parallel, a similar development took place at other places and in other federal states.

In this phase, the achievable accuracy and the economic efficiency of cadastral photogrammetry compared to conventional methods were discussed. The advantages of cadastral photogrammetry could be exploited especially in the application in land consolidation areas. As control points,

mostly survey points from the already existing official fixed point field could be used. In addition, a large number of new points had to be determined in the land consolidation areas (KRAUS 1973). In the meantime, the availability and the capacities of the computer systems had increased considerably. Thus, especially the program developments were advanced. At the Institute of Photogrammetry, the programs PAT-M-4 (position adjustment) and subsequently PAT-M-43 (spatial adjustment) were developed, which found application especially in cadastral photogrammetry. Both programs were used for block adjustment according to independent models. In parallel, other methods and programs were developed which were intended to increase the accuracy of point determination (ACKERMANN 1973).

In addition to the block adjustment according to independent models, the method of bundle block adjustment was already investigated at that time. For the adjustment according to independent models, however, more evaluation devices were available in practice and the bundle block adjustment proved to be not as accurate as expected for the time being. Thus, adjustment by independent models was preferred and became widespread in practical applications. Bundle block adjustment was not to become relevant again until a decade later (ACKERMANN 1973).

In January 1971, a course on numerical photogrammetry was held at the Technical Academy in Esslingen. Theoretical and practical contents, as well as the developed software system were presented to the public. In further publications, the potential of numerical photogrammetry was recognized and block adjustment and point determination were judged to be accurate and reliable. As a result, cadastral photogrammetry was approved as a method for determining point coordinates in October 1971. This was realized by the administrative regulation VVPhoto and the technical guideline TRPhoto, which came into force on October 14, 1971. Since then, cadastral photogrammetry has been used regularly for land consolidation or road closure surveys.

Between 1970 and 1973, photogrammetric point determination by block adjustment was widely applied and tested. A large proportion of these projects involved practical applications in cadastral photogrammetry for land consolidation procedures. Thus, the responsible State Office for Land Consolidation and Settlement had firmly introduced appropriate evaluation equipment and the photogrammetric block method. At that time, the state of numerical photogrammetry and the block adjustment used according to independent models were considered successful and particularly reliable (ACKERMANN 1985).

For the application in the real estate cadastre, the double survey or a hybrid solution was available for selection. In the hybrid solution, the coordinates are created from photogrammetric and terrestrial observations. The decision was made in favour of the hybrid solution, as this was expected to result in higher neighbourhood fidelity. In this case, span dimensions as well as other geometric conditions were introduced into the adjustment, which were recorded terrestrially between the signalized new points. The so-called "span adjustment" was able to establish the neighbourhood fidelity within the area, which is important in the real estate cadastre. Another advantage was that the provisional coordinates could be checked and gross errors, for example in signalization, could be detected. During the terrestrial recording of the span dimensions, points that were not visible from the air could be recorded at the same time. The terrestrially recorded elements made it possible to determine these points later by calculation. In addition to span

adjustment, least squares interpolation was introduced to increase accuracy. This reduces network stresses and the residual errors at the control points (Ackermann 1973).

The period between 1970 and 1973 can be considered as the development and introduction phase. Between 1970 and 1983, cadastral photogrammetry has been applied in 142 cadastral surveying procedures in Baden-Württemberg. In this operational phase the development of the existing methods was reviewed and advanced (ACKERMANN 1985; WALDBAUER 1985a; WALDBAUER 1985b).

After 1983: At that time, it became known that clear tensions exist in the procedures from the seventies. Accordingly, the block adjustment according to independent models yielded very good internal accuracies, but did not correspond to the theoretical accuracies due to systematic image errors. From 1983 onwards, the introduction of bundle block adjustment and improvements of the evaluation equipment demonstrably improved the accuracy significantly. From then on, model coordinates were no longer measured. Instead, the modern analytical evaluation instrument Planicom C100 was available for the measurement of image coordinates (ACKERMANN 1985).

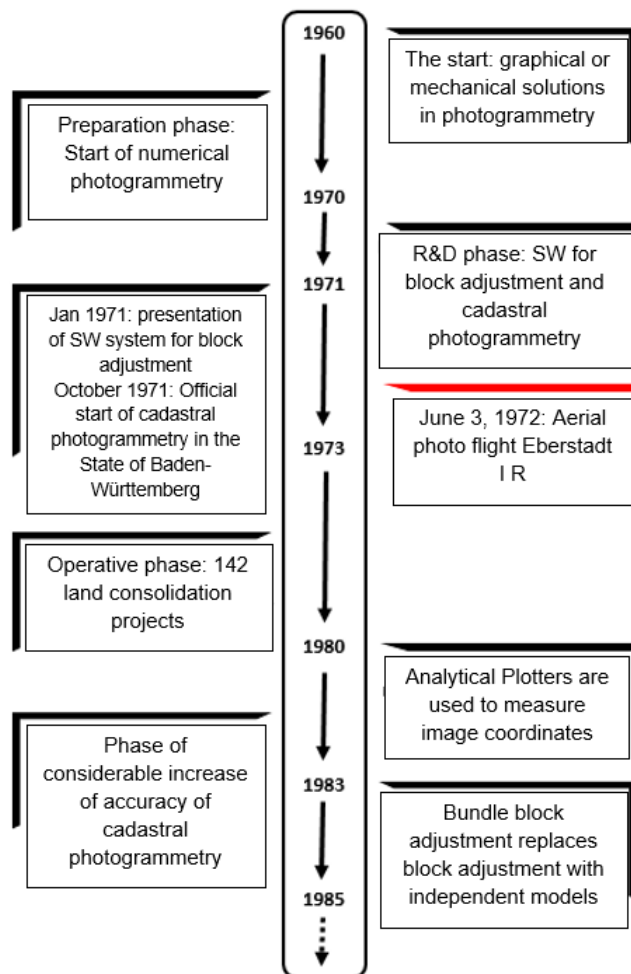


Fig. 1: Timeline of development of Cadastral Photogrammetry in the State of Baden-Württemberg

### 3 Land consolidation project Eberstadt I R

#### 3.1 Area description

The former vineyard consolidation area Eberstadt I R is located in the district of Heilbronn north of the municipality of Eberstadt. The abbreviation "I R" stands for the first vineyard consolidation procedure in this district.

The Eberstadt I R land consolidation procedure was ordered in 1967. The area of the land consolidation is 181 ha, of which 113 ha are forest and 44 ha are vineyards. In 1977, the final determination of the land consolidation took place.

In order to transfer the coordinates of the new boundaries into the real estate cadastre, the method of cadastral photogrammetry was chosen in this land consolidation procedure and an aerial photo flight was carried out to determine the coordinates. However, no information is available on the aerial photo evaluation, block adjustment and point determination at that time. As described in chapter 2, however, it can be assumed on the basis of the image flight date that the block adjustment was used according to the method of independent models.

#### 3.2 Landslide

Landslides, as they occur at the Eberfirst, are known in the whole Württemberg Keuperbergland, especially east of the Neckar. Responsible for the landslides at the Eberfirst are glacial large-scale landslides which occur in the border area between gypsum keuper and reed sandstone. Thereby, a smooth landslide surface is present between two different geological layers. Glacial large-scale landslides are actually old, closed landslides that can be reactivated by human intervention and/or by special climatic conditions. On the Eberfirst, soil relocation took place as part of the 1970 vineyard consolidation, which included filling and levelling many blades. Due to this intervention and heavy rainfall, the slope began to slide again, (LEO-BW 2022; LGRB 2022a; LGRB 2022b).



Fig. 2: Visible break-off edge (Drone image by Heilbronn Surveying office)

The landslides are limited to the western part of the Eberfirst vineyard. The earth movements are not only known to the surveyor's office, but also take on dimensions that make them noticeable to property owners and residents. These are noticeable by cracks and unevenness in the asphalt of the paths or by visible break-off edges in the terrain and affect the cultivators of the vineyards (Figure 2).



### 3.3 Historical aerial photo flight

The aerial photo flight in the land consolidation procedure Eberstadt I R was carried out on June 3, 1972. In order to cover the area sufficiently, 16 aerial images were taken in one flight strip. The individual aerial images of the flight strip have a coverage of about 80 percent. The flight altitude of the aerial survey is given as 760 m above ground, resulting in an image scale of 1:5000. The image flight was carried out by the company Geoplana with a film-camera RC8 of the manufacturer Wild Heerbrugg. The lens used has a camera constant of 152.55 mm. The camera calibration protocol of May 31, 1972 is known, so that the calibration values can be used for the new evaluation. Figure 3 shows an aerial image from the Eberstadt I R image flight. At that time, the aerial photos were exposed in the standardized format of  $23 \times 23$  cm on black-and-white film rolls.

However, as an additional difficulty in the evaluation of the land consolidation procedure Eberstadt I R the signalization sketches are no longer available. This means a more difficult assignment of the signalized points in the aerial images, as well as missing span dimensions for the control of the generated coordinates.



Fig. 3: Aerial image from the aerial photo flight Eberstadt I R scanned at Heilbronn Surveying office

## 4 Workflows for re-evaluation of historical aerial images

Two basic workflows have been used for the re-evaluation: a currently developed workflow for cadastral photogrammetry and a workflow for the detection of landslides. The basic features are presented here and the empirical tests in the Eberstadt I R procedure are evaluated in chapter 5.

### 4.1 Workflow Cadastral Photogrammetry

In order to re-evaluate the former land consolidation areas using cadastral photogrammetry methods, a uniform workflow is currently developed at the Heilbronn survey office. The

processing is divided into several steps. Based on a thorough project preparation and the digitization of the analogue aerial images, suitable control points have to be selected and identified and measured in the field. After an interior orientation of the images using the camera calibration information and an aerial triangulation the basis is prepared for the point determination of ground control-, boundary- and polygon points. These points have to be checked to be finally transferred to the real estate cadastre.

## 4.2 New aerial triangulation for project Eberstadt I R

As a basis for the landslide detection according to chapter 4.3 the results of the new aerial triangulation for project Eberstadt I R are described below.

### 4.2.1 Scanning of aerial images

The aerial images have been digitized with an UltraScan5000 of Vexcel with a pixel size of  $10 \mu\text{m} \times 10 \mu\text{m}$ .

### 4.2.2 Measurement of control points

For the necessary control points for the photogrammetric evaluation, boundary and polygon points have been recorded in Eberstadt on three field service days. With the previous UTM-G coordinates the points have been staked out and uncovered. The control points are chosen to be evenly distributed in and around the former land consolidation area. No control points were recorded in the suspected landslide area, since there is no longer any positional identity there. A total of 36 ground control points have been recorded, of which 20 are polygon points and 16 are boundary points (Figure 4).

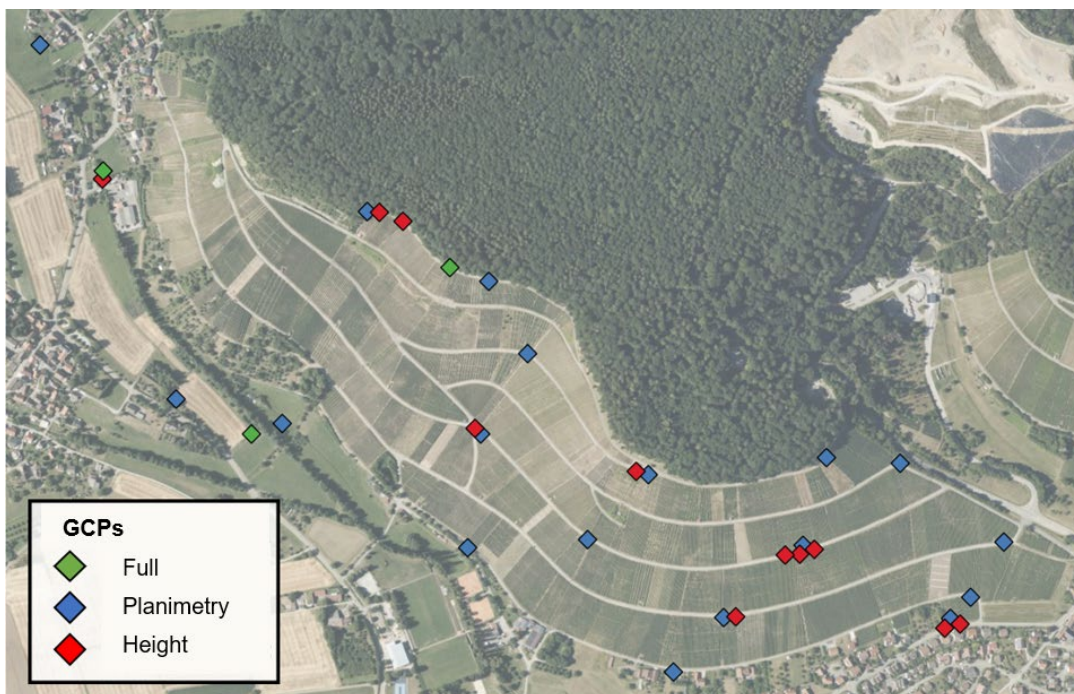


Fig. 4: Distribution of ground control points

### 4.2.3 Aerial triangulation

The RMSE (Root-mean-square deviation) value for the fiducial measurements was in the range of 0.3 to 0.7 pixel. 476 tie points were used. A total of 33 GCP (Ground Control Points) were used. Of these, 3 can be used as full, 18 as planimetric and 12 as height GCP. The distribution of control points is shown in Figure 4, with the control point types distinguished by color. The RMSE values in X/Y/Z for the ground control points were 0.0344 m / 0.0274m / 0.0282 m respectively. The image measurement accuracy at GCPs was for x/y 0.2956 pixel /0.3297 pixel respectively. During the photogrammetric point determination, new coordinates were calculated for 1326 boundary and polygon points. On average, a point was determined with five individual measurements.

### 4.3 Workflow for detecting landslides

The basic idea of detecting landslides is to compare a current and a historical terrain model. If the terrain has changed in the meantime, for example due to landslides, this change will show up when comparing the two terrain models. The current terrain is represented by the digital terrain model DTM1 of the national survey. The historical terrain model is derived from the available aerial images and represents the condition in 1972. For this purpose, the workflow shown in Figure 5 was developed, which consists of seven processing steps. The initial situation is that the historical images are oriented by means of an aerial triangulation. A point cloud is derived from the historical aerial images and is further processed with classifications and filtering. If the processed point cloud contains only the relevant ground points, it can be considered as a historical terrain model. The Erdas Imagine and CloudCompare software is used to perform the developed workflow.

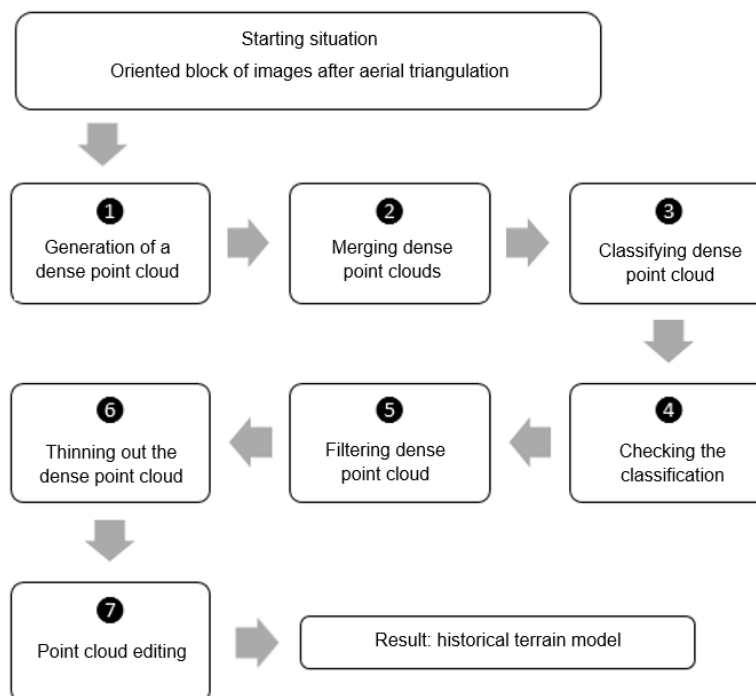


Fig. 5: Workflow for the derivation of the historical terrain model

## 5 Empirical tests and evaluations in the land consolidation procedure Eberstadt I R

### 5.1 Accuracy and control

Table 1 compares parameters of the aerial triangulation of four different re-evaluations of former land consolidation procedures. All four test procedures have been evaluated by the surveyor's office Heilbronn with the same workflow from chapter 4. The compared residual errors at the control points are in a similar spectrum. The procedure Eberstadt I R, which is affected by landslides, does not have higher residual errors at the control points too.

It is important to note that there must be a sufficient occupation of planimetric and/or full control points, especially at block edges (RAHN 2021).

Tab. 1: Comparison of aerial triangulation results with 3 other land consolidation projects

Procedure		Beilstein - Etzlenswenden	Flein - Schenkenhalde	Nordheim - IX R	Eberstadt - I R
Residuals [m]	X	0,0214	0,0273	0,0323	0,0344
	Y	0,0201	0,0278	0,0456	0,0274
	Z	0,0195	0,0390	0,0441	0,0282
Total RMSE [Pixel]		0,4771	0,4110	0,4588	0,3859

With a complex adjustment the relative position of the generated coordinates can be controlled. The Soldner coordinates of the former determination and the UTM coordinates of the new determination are compared. As a result of the statistical check a protocol and an overview plan are generated. In the protocol UTM and Soldner coordinates of identical points are compared. Possible gross errors are statistically indicated in the protocol. These possible gross errors are also displayed in the overview plan. If a point vector is conspicuously long and at the same time points in a different direction than the neighboring point vectors, there is a possible gross error. The gross error may have occurred during the determination at that time or during the determination today. In case of a possible gross error, further investigations are therefore necessary. The gross error can be narrowed down by a GNSS control survey or by a span control.

If the vectors point locally (along the agricultural road) in the same direction and have a similar length as shown in Figure 6, a homogeneous determination is present. For the most part, the newly determined points in the Eberstadt I R procedure have been determined with good neighbourhood fidelity. In the area of the landslide no different behaviour is recognizable between neighbouring points. By checking with the complex adjustment, however, no statements can be made about the accuracy of the absolute point position.

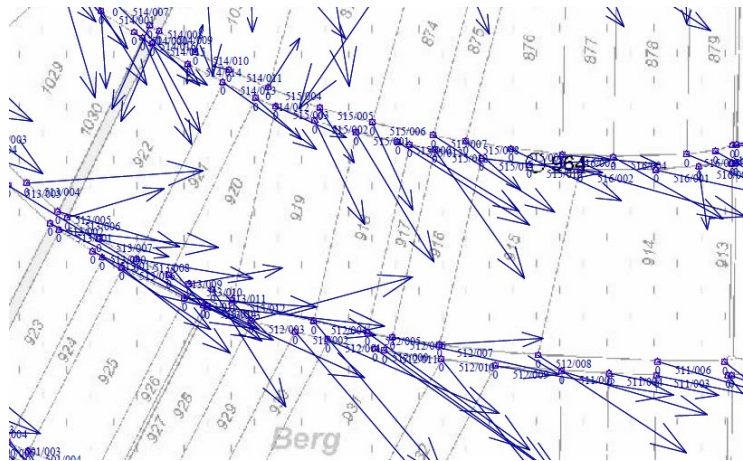


Fig. 6: Part of the result of the complex adjustment result with point vectors

After coordinates for the new points have been generated by photogrammetry, these are to be checked by means of random samples in the field comparison. For this purpose, the generated coordinates are loaded into DAVID-kaRIBik. In the field, a boundary check of the corresponding boundary point is performed with a GNSS survey. With this approach 47 boundary points have been recorded over the area. After evaluation of the measurement data, the positional difference between the coordinates from the redetermination and the coordinates of the actual position can be calculated.

According to the current administrative regulation for the execution of real estate surveys (VwVLV 2022), state coordinates of a boundary point may have a linear deviation of 8 cm when verified. At the time of the survey in 1972, the Administrative Regulations on Error Limits for Cadastral Surveys (VVFEHLERGRENZEN 1971) applied. According to this, a maximum deviation of 8 cm was permissible when checking a boundary point. The extent to which the regulation of that time can also be applied to today's evaluation has not been conclusively clarified.

Of the 47 boundary points checked in a first investigation 30 points possess a deviation of better than 8 cm, which is 64 percent. It should be mentioned here that better values were seen in previous test procedures.

It is noticeable that the Eberstadt I R procedure is larger than the previous procedures and that significantly more new points were determined. It could therefore be possible that in a larger area increased stresses occur, which cannot be completely compensated by the aerial triangulation. Thus, remaining residual errors could possibly also have a negative effect on the coordinates of the redetermination. In addition, it must be noted that the Eberstadt I R procedure was flown and evaluated in 1972. The procedures of Nordheim - IX R and Flein – Schenkenhalde were carried out much later. As described in chapter 2, cadastral photogrammetry was not fully developed at the beginning and the accuracy achieved in practice was not significantly increased until 1980. Thus, it is conceivable that there are systematic errors in the methods and data basis of that time that remain in the coordinates of today's evaluation.

## 5.2 Landslide area

By using the workflow developed in chapter 4.3, a clear landslide area could be detected. The detected landslide area also coincides with terrain changes visible on site. As assumed, this area is located in the western part of the land consolidation procedure and starts with a clear break-off edge, which is located about 272 m above sea level. This is located along the red colouring in Figure 7, is approximately 280 meters long, and is also clearly visible in the terrain. In the area of the break-off edge, the current terrain has subsided by three to five meters compared to its condition in 1972. Below the break-off edge is a 30 to 40 meter wide strip that has slumped one to two meters due to the landslide. This is followed by a transition area where no elevation change was detected. Below this is the area coloured blue in Figure 7. Here, too, a landslide is present, whereby the current ground level is about one meter higher than the original terrain. The volume of soil that has slipped in the area of the break-off edge thus accumulates again further down the slope. In total, a landslide area of approximately 87,500 m<sup>2</sup> can be clearly detected with the developed workflow.

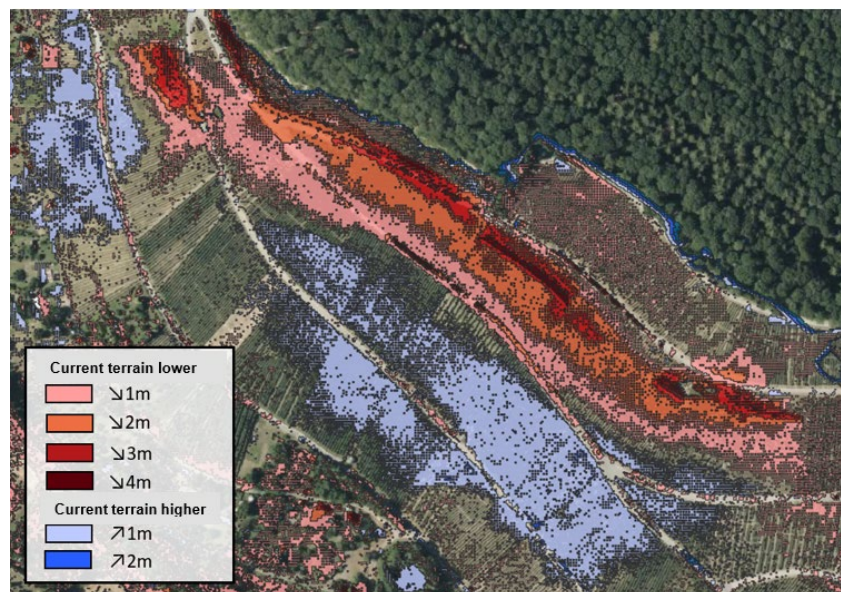


Fig. 7: Detected landslide area and changes in heights compared to 1972

In order to compare the absolute terrain heights of today and back then, it is suitable to compare them in a terrain profile. As an example, a terrain profile is shown in Figure 8. The selected area is particularly suitable for comparison because significant terrain changes are assumed locally there. The terrain changes can therefore also be confirmed by a terrain profile as visible in Figure 8. The original terrain of 1972 (green) has a uniform slope. Between 70 and 80 meters, the slope of the terrain deviates because a path is located there. In comparison, today's terrain (pink) is not so uniform. Within the first 50 meters, the present terrain is clearly indented and is below the original terrain. In this slide area, the terrain has slumped up to three meters in elevation. In the subsequent area, the present terrain is mostly one to two meters above the original terrain. It is also visible that the named road is now located further downslope and must also have been affected by the landslide.

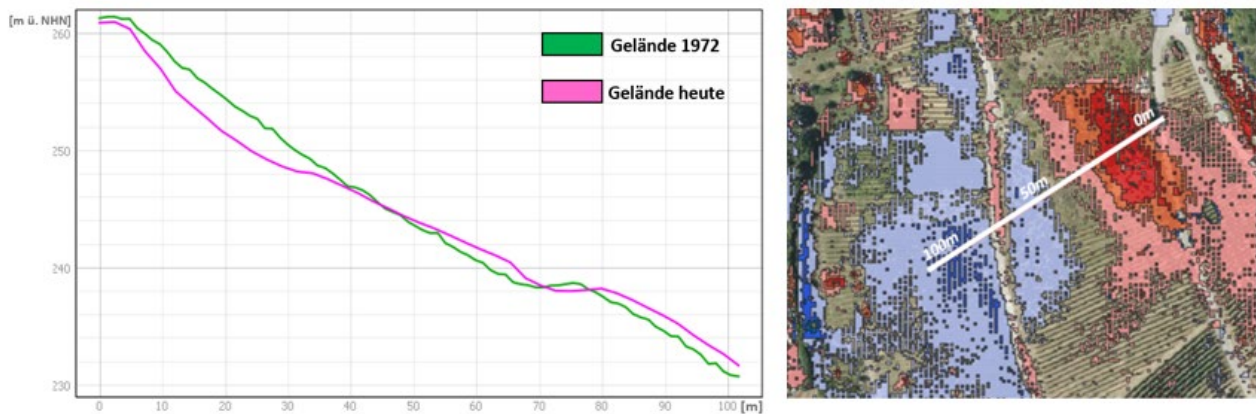


Fig. 8: Cross-section profile of landslide

The developed workflow can only detect landslides that are related to a change in elevation in the terrain. However, landslides can also occur without a change in the elevation of the terrain. For this, the terrain must have a uniform slope and the landslide must not deform the terrain at the surface. If no changes in elevation and terrain deformation occur in the landslide, there will be no changes in the digital terrain model. Therefore, in this case, landslide cannot be detected by the developed workflow. Figure 9 sketches this situation, where the landslide can be detected in use case 1. In use case 2, the landslide cannot be detected by the developed workflow, which was partly the case in the project area. It was possible to precisely delimit an 87,500 m<sup>2</sup> landslide area with changes in elevation. A maximal height difference of 5 m and a maximal planimetric change of 2.7 m could be found.

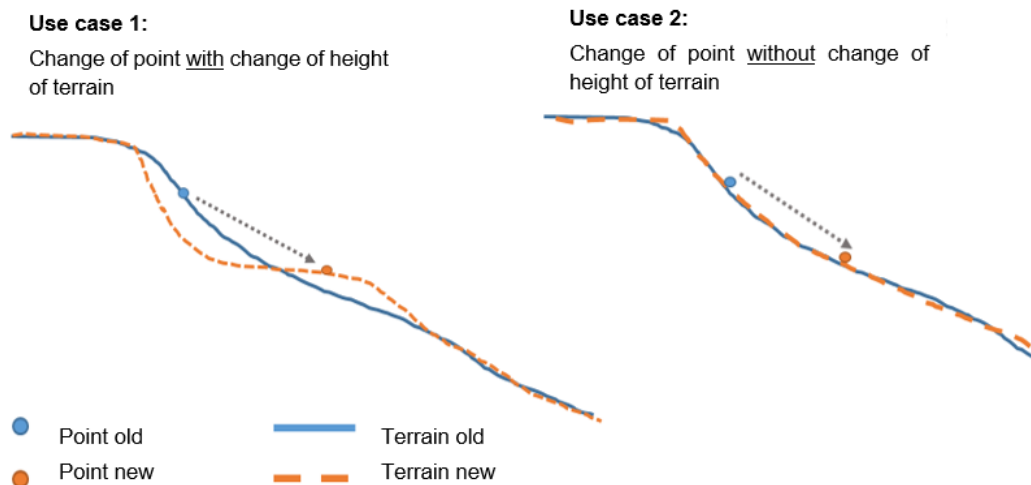


Fig. 9: Two use cases for landslides. Case 1 can be determined by the developed procedure

By comparing the orthophoto with the boundaries of the property map or by comparing site coordinates, the landslide can nevertheless be delineated fairly accurately. The total area of the landslide is approximately 141,000 m<sup>2</sup>, which is approximately 32 percent of the total area.

## 6 Conclusions

In conclusion, it can be stated that the photogrammetric re-evaluation of the former land consolidation area could be carried out successfully after exactly 50 years. The workflow being developed by the surveyor's office Heilbronn could be applied despite the existing landslide area. The aerial triangulation could be carried out successfully during the re-evaluation. A sufficient number of control points were available, which could also cover the landslide area. The residual errors at the control points used also remained within the expected range. They are comparable with the empirical values from previous test procedures.

In a complex adjustment, the Soldner coordinates of the former determination have been compared with the UTM coordinates of the present determination. This is to check the homogeneity of the photogrammetric determination. During the comparison of the coordinates a neighbourhood-faithful determination was predominantly proven along the ways, whereby outliers were detected only sporadically.

A workflow could be developed that can detect landslides by changes in the terrain. This involves comparing a current terrain model with a historical terrain model. The most current terrain model can be obtained from the State Office for Geoinformation and Land Development. The historical terrain model is generated from the historical aerial images after aerotriangulation has been performed. A point cloud is generated from the historical aerial images, which is then classified and filtered. If the processed point cloud contains only the relevant ground points it can be considered as a historical terrain model. When comparing the two terrain models, elevation differences can be coloured by area and thus the landslide can be visualized. The landslide can also be viewed in a terrain profile. In the Eberstadt I R procedure, it was thus possible to locate a clear break-off edge of the landslide. It was also possible to precisely delimit an 87,500 m<sup>2</sup> landslide area with changes in elevation. In the Eberstadt I R procedure, however, there are also landslides without significant changes in elevation, which can only be detected by other methods. The comparison of the orthophoto with the boundaries of the real estate cadastre or the comparison of position coordinates can complement the developed workflow in the detection of landslides.

## 7 Acknowledgements

This paper is dedicated to late Prof. Dr. Ackermann, the leading photogrammetrist of the 20th century who was a pioneer in cadastral photogrammetry, aerial triangulation, image matching and many other topics. The paper is based on the developments and findings of the Bachelor-Thesis by B.Eng. Adrian Weng, which was conducted in winter term 2021/22 at the Landratsamt Heilbronn, Vermessungsamt and supervised by Prof. Dr.-Ing. Eberhard Gülch, HFT Stuttgart and Dr.-Ing. Franziska Wild, Landratsamt Heilbronn, Vermessungsamt.



## 8 References

- ACKERMANN, F., 1973: Numerische Photogrammetrie - Zwischenbilanz einer intensiven Entwicklung. Numerische Photogrammetrie, Herbert Wichmann Verlag, Karlsruhe, 1-43.
- ACKERMANN, F., 1985: Photogrammetrische Punktbestimmung. Sonderheft zur Katasterphotogrammetrie der DVW-Mitteilungen des Landesverein Baden-Württemberg, 7-34.
- KRAUS, K., 1973: Numerische Photogrammetrie - Zwischenbilanz einer intensiven Entwicklung. In F. Ackermann, Numerische Photogrammetrie, Herbert Wichmann Verlag, Karlsruhe, 45-79.
- LEO-BW, 2022: LEO-BW (landeskundliche Informationssystem für Baden-Württemberg) - Detailseite Eberstadt. [https://www.leo-bw.de/en\\_US/web/guest/detail-gis/-/Detail/details/ORT/labw\\_ortslexikon/1904/Eberstadt](https://www.leo-bw.de/en_US/web/guest/detail-gis/-/Detail/details/ORT/labw_ortslexikon/1904/Eberstadt), last access 03.01.2023.
- LGRB, 2022a: LGRBwissen (Landesamt für Geologie, Rohstoffe und Bergbau) - Glossar. <https://lgrbwissen.lgrb-bw.de/glossar>, last access 04.01.2023.
- LGRB, 2022b: LGRBwissen (Landesamt für Geologie, Rohstoffe und Bergbau) - Hangrutschung „Wildenberger Hang“. <https://lgrbwissen.lgrb-bw.de/ingenieurgeologie/massenbewegungen/hangrutschungen-rutschen/hangrutschungen-den-gesteinen-des-mittelkeupers-grenzbereich-grabfeld-formation-zur-stuttgart-formation/hangrutschung-wildenberger-hang>, last access 04.01.2023.
- RAHN, A., 2021: Auswertung historischer Luftbilder aus Flurbereinigungsverfahren. Stuttgart: Institut für Photogrammetrie der Universität Stuttgart.
- VVFEHLERGRENZEN, 1971: Verwaltungsvorschrift über die Fehlergrenzen bei Katastervermessungen (VV Fehlergrenzen). Nr. II 2 2003/93, 20.07.1971.
- VwVLV, 2022: Verwaltungsvorschrift des Ministeriums für Landesentwicklung und Wohnen für die Durchführung von Liegenschaftsvermessungen (LV-Vorschrift – VwVLV), 12. April 2022 – Az. MLW16-28-161/1.
- WALDBAUER, G., 1985a: Blockausgleichung mit unabhängigen Modellen bei der photogrammetrischen Katastervermessung der Flurbereinigungsverwaltung zwischen 1970 und 1983. Sonderheft zur Katasterphotogrammetrie der DVW-Mitteilungen des Landesverein Baden-Württemberg., 35-53.
- WALDBAUER, G., 1985b: Bündelblockausgleichung im praktischen Einsatz bei der Flurbereinigungsverwaltung. Sonderheft zur Katasterphotogrammetrie der DVW-Mitteilungen des Landesverein Baden-Württemberg., 84-94.
- WILD-PFEIFFER, F., KAISER, F., STEPAN, D., FITTINGER, R. & WEISER, G., 2018: Katasterphotogrammetrie – besser als ihr Ruf? Mitteilungen und Veröffentlichungen zum Vermessungswesen aus den Themenbereichen Geodäsie, Geoinformation und Landmanagement / DVW Landesverein Baden-Württemberg., Heft 1/2018, DVW-Landesverein Baden-Württemberg, Vorstand (Hrsg.), 5-24.

# Simulation of Earth Observation Data Utilizing a Virtual Satellite Camera

NINO LENZ<sup>1</sup> & MICHAEL GREZA<sup>1</sup>

*Abstract: Image chain simulation is an essential tool for the design of Earth observation satellites. In this work, a simulator for the images produced by an upcoming low Earth orbit remote sensing nanosatellite constellation by the Technical University of Munich, Center for Telematics Würzburg and the Bavarian Surveying and Mapping Authority is developed. Projection and sampling models adjusted for this application are presented. The system's modulation transfer function is estimated according to analytical models for atmosphere, sensor, and optical system effects. Images are simulated based on digital orthophotos and the desired satellite orbit.*

*Disturbances of the ideal orbits are introduced, and their impact on the appearance of the produced images is explored. A comparison with other nanosatellite remote sensing missions is used to estimate expected inaccuracies resulting from attitude deviations. The localization of real and fictional ground control points is employed to estimate positional inaccuracies in the simulated images caused by a lack of information about terrain shape and elevation.*

## 1 Introduction

Historically, satellite missions have been characterized by long development cycles and high costs, in part due to the large size of satellites. However, technical advances have made much smaller designs feasible in the last few decades. Standardized dimensions allow the deployment of many nanosatellites at once by hitching a ride on a larger launch vehicle as a secondary payload. Economies of scale and reduced launch costs have dramatically lowered the entry barrier to satellite missions (CRUSAN & GALICA 2019).

Unlike many other engineering applications, for satellites, adjustments to the system design can - if at all - only be made in software after deployment, and serviceability is essentially zero. Where functionality cannot be tested directly, extensive simulation of the system is therefore required to assess the expected behavior. A simulation of the imaging subsystem can confirm that goals and requirements of the mission will be met or indicate if design changes need to be made.

In this work, a simulator for the images produced by a low Earth orbit remote sensing nanosatellite with a ground sampling distance (GSD) of roughly 4 m is presented. A priority is the precise rendering of the planned satellite orbit and the resulting ground coverage. In addition, the appearance of the produced images is adjusted to account for properties of a line scan camera, atmosphere, and satellite movement. The focus will be on questions of spatial resolution, rather than radiometric qualities. Several simulations are run to assess the expected image quality, and to investigate the influence of different parameters on the simulation results.

---

<sup>1</sup> Technische Universität München, Photogrammetrie und Fernerkundung, Arcisstr. 21, 80333 München, E-Mail: [nino.lenz, michael.greza]@tum.de

The generation of artificial satellite imagery is an experimental part of a simulation toolkit in an ongoing research collaboration for the development of a low Earth orbit remote sensing nanosatellite constellation between the Technical University of Munich, the Center for Telematics Würzburg and the Bavarian Surveying and Mapping Authority.

## 2 Related Work

Previous work on the simulation of satellite-based Earth-observation data tends to focus more on the resulting image quality with regards to MTF, resolution, noise, or radiometric characteristics, rather than the precise simulation of viewing geometry and position based on prescribed orbits. In accordance with this, several authors chose to carry out most calculations in the frequency domain (e.g. ALICI et al. 2019; COTA et al. 2008; WAHBALLAH et al. 2021). However, since certain properties of imaging systems are represented in the spatial domain more easily or precisely, some research has been done in this regard as well (BÖRNER et al. 2001; CHEN et al. 2009). Regardless of spatial or frequency domain, the simulations tend to follow a similar scheme.

The input dataset should provide significantly higher spatial and/or spectral resolution than the simulated imaging system and high SNR to be suitable for use as “ground truth” data, on which the simulation is based. The use of approximate inferences about the acquisition conditions, in conjunction with other software tools such as MODTRAN, allows for the conversion of aerial imagery data to values of bottom-of-atmosphere radiance.

In subsequent steps, effects representing illumination conditions and interactions with the atmosphere, the camera’s optical system, movement of the satellite, and properties of the camera sensor are applied. The modified input data is downsampled and quantized to receive the output image.

## 3 Methods

### 3.1 Input Data

The input data for the simulation consists of the nominal satellite orbits, containing satellite position and attitude, a digital orthophoto (DOP) of the region to be recorded, and a list of ground control points (GCPs) whose position in the output image is to be determined. The orbits are provided by Astronomical and Physical Geodesy of TU Munich while the DOPs and GCPs are provided by the Bavarian Surveying and Mapping Authority.

In the context of the simulator, the DOP is assumed as the ground truth, i.e., as an image obtained by an ideal camera, without any losses in resolution or other effects degrading the image quality. In reality, the DOP was recorded using real, imperfect cameras. However, since the expected resolution of the simulated images (GSD  $\sim 4$  m) is much lower than that of the DOP (0.2 – 0.4 m), these imperfections are regarded to have negligible impact on the simulation results.

### 3.2 Simulator Design Procedure

To model the degradation of the data due to the simulated imaging process, a system model is created, and supplied with mission-specific camera and orbit parameters. Steps in the image chain

are simulated individually, and finally compiled into an overall modulation transfer function (MTF). Where detailed mission-specific information is not yet available, typical values are substituted from literature.

It should be noted that only narrow-bandwidth spectra will be considered here, such that all wavelength-dependent quantities can be assumed as constant within a single band. For sensors with multiple narrow bands, the wavelength-dependent effects should therefore be applied to each band separately. While the DOP used in this paper provides only RGB bands, the simulator is capable of generating multispectral images as well, given suitable input data.

Further, a sophisticated attempt to determine radiometric properties, as mentioned in section 2, is omitted due to its reliance on external software and since questions of spectral resolution are not the focus of this work. In absence of more information about the original acquisition system, the DOP data is instead assumed to express bottom-of-atmosphere radiance values of unknown units. As such, a high degree of radiometric accuracy cannot be expected from the simulated images.

### **3.3 Experiments**

Once suitable parameters for the simulation have been determined, the image swaths are simulated according to the provided nominal satellite orbits. The system MTF curve and its components are presented.

In addition, the effects of deviations from nominal attitude are investigated. For this purpose, sinusoidal oscillations in the satellite's pitch and roll angles are introduced to simulate gyroscopic motion. The magnitude and frequency of the deviations are varied in order to estimate their effects on image quality. Allowable tolerances and different effects on the appearance of produced images are explored.

The impact of elevation differences on positional accuracy is investigated. To this end, the positions of GCPs are specified once at their actual elevations, and once assumed to all lie in the same horizontal plane. The differences in their positions in the simulated images are determined to assess the possible benefits of introducing a digital surface model (DSM). In a second step, fictional GCPs are introduced to better quantify the errors caused by the absence of a DSM. Determining parameters for the magnitude of the errors are identified and a model for rough error estimation presented.

## **4 Simulator Design**

### **4.1 Simulation Workflow**

To satisfy the requirement of high positional fidelity, the primary part of the image acquisition is simulated in the spatial domain. According to the projection model discussed below, the image is assembled line-by-line based on the given satellite positions and attitudes. In a second step, this "raw" simulated image is degraded further by applying the MTF and finally adding noise.

### **4.2 Projection Model**

The line sensor is continuously exposed to the incident radiation, and the accumulated charge read out in regular intervals. These aspects are taken into account by the following model.

The terrain, or rather the input DOP, is assumed to be perfectly flat and lie in a horizontal plane at a specified elevation. This elevation should therefore be chosen to represent the region of interest as well as possible.

The line sensor is modelled as a one-dimensional line, which is made up of segments representing an individual pixel each. The pixels' corner points are then projected onto the DOP plane. This projection is performed once from the exterior orientation at the beginning and again at the end of each sampling period. By constructing quadrilaterals from the four corner coordinates obtained for each pixel in this way, the continuous area covered by the satellite swath is obtained, without the possibility of gaps occurring between subsequent exposures. Figure 1 illustrates this projection model.

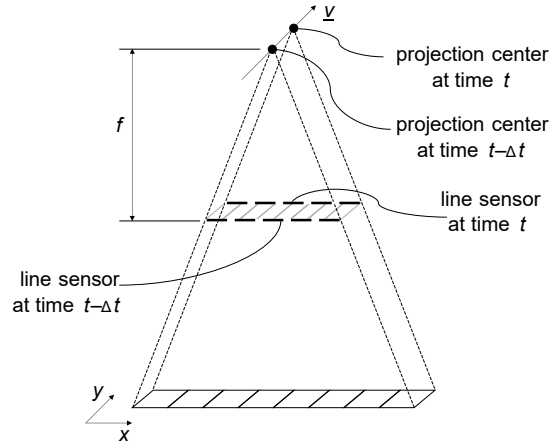


Fig. 1: Projection model

Once the projected quadrilateral corresponding to a pixel has been obtained in this way, the given GCPs are checked for intersection with the projected area using a ray tracing algorithm to ensure correct localization independent of a GCP's elevation. Finally, the value of the pixel in the raw simulated image is computed as the mean value of all DOP pixels contained within the quadrilateral.

The resulting projection is thus a mix of the central and parallel projection models: While it realizes a central perspective in the across-track direction, the along-track sweep is modelled in parallel perspective.

### 4.3 Modulation Transfer Function

The individual steps in the image chain are simulated by choosing appropriate analytical MTF models, and parameters are adjusted based on the camera datasheet, orbit data, and typical values found in literature. The MTF is expressed in terms of spatial frequencies in the focal plane, with  $k_x$  denoting frequencies in the across-swath direction, and  $k_y$  denoting frequencies along the flight direction. The following definitions are useful for discussion of some of the imaging steps: the radial spatial frequency  $k_r$ , diffraction-limited cutoff frequency  $k_{co}$ , and normalized spatial frequency  $k_n$ , as given in eq. 1.

$$k_r = \sqrt{k_x^2 + k_y^2}, \quad k_{co} = \frac{D}{\lambda f}, \quad k_n = \frac{k_r}{k_{co}} \quad (1)$$

with primary aperture diameter  $D$ , incoming radiation wavelength  $\lambda$ , and focal length  $f$ .

Turbulence in the atmosphere degrades image quality according to eq. 2 to 3 (ALICI et al. 2019; HOLST 2000), while aerosols cause scattering and absorption according to eq. 4 (ALICI et al. 2019; HOLST 2000; SADOT & KOPEIKA 1994).

$$MTF_{turbulence} = \exp\left(-3.44 \left(\frac{\lambda f k_r}{r_0}\right)^{\frac{5}{3}} \left[1 - \alpha \left(\frac{\lambda f k_r}{D}\right)^{\frac{1}{3}}\right]\right) \quad (2)$$

$$r_0 = 0.185 \lambda^{\frac{6}{5}} \left[ \int_0^H \left(\frac{h}{H}\right)^{\frac{5}{3}} C_n^2(h) dh \right]^{-\frac{3}{5}}, \quad C_n^2(h) = 4.16 \times 10^{-13} m^{\frac{2}{3}} h^{-\frac{4}{3}} \quad (3)$$

where  $r_0$  is the atmospheric coherence diameter, and  $H$  the satellite's altitude above ground.

$$MTF_{aerosol} = \begin{cases} \exp\left(-A_a l_{opt} - S_a l_{opt} \left(\frac{k_r}{k_{aco}}\right)^2\right), & \frac{k_r}{k_{aco}} \leq 1 \\ \exp\left(-A_a l_{opt} - S_a l_{opt}\right), & \frac{k_r}{k_{aco}} > 1 \end{cases} \quad (4)$$

with atmospheric absorption and scattering coefficients  $A_a$  and  $S_a$ , optical path length  $l_{opt}$ , and aerosol cutoff frequency  $k_{aco}$ .

Diffraction around its aperture limits the spatial frequencies resolvable by the camera. A suitable model for the simulated Cassegrain optical system is used to represent the decreasing MTF behavior (COTA et al. 2008; O'NEILL 1956), as shown in eq. 5 through 9.

$$MTF_{diffraction} = \frac{2}{\pi} \frac{A+B+C}{1-\epsilon^2} \quad (5)$$

$$A = \arccos(k_n) - k_n \sqrt{1 - k_n^2} \quad (6)$$

$$B = \begin{cases} \epsilon^2 \left[ \arccos\left(\frac{k_n}{\epsilon}\right) - \frac{k_n}{\epsilon} \sqrt{1 - \left(\frac{k_n}{\epsilon}\right)^2} \right], & k_n \leq \epsilon \\ 0, & k_n > \epsilon \end{cases} \quad (7)$$

$$C = \begin{cases} -\pi\epsilon^2, & k_n \leq \frac{1-\epsilon}{2} \\ -\pi\epsilon^2 + \epsilon \sin(\phi) + \frac{\phi}{2} (1 + \epsilon^2) - (1 - \epsilon^2) \arctan\left(\frac{1+\epsilon}{1-\epsilon} \tan\left(\frac{\phi}{2}\right)\right), & \frac{1-\epsilon}{2} < k_n \leq \frac{1+\epsilon}{2} \\ 0, & k_n > \frac{1+\epsilon}{2} \end{cases} \quad (8)$$

$$\phi = \arccos\left(\frac{1+\epsilon^2-4k_n^2}{2\epsilon}\right), \quad \epsilon = \frac{D_{obscuration}}{D} \quad (9)$$

with  $D$  the primary aperture diameter, and  $D_{obscuration}$  the diameter of the circular obscuration mirror.

In addition, a variety of aberrations in the optical system can affect image quality. Since their exact nature is generally not known in the early stages of the design process, their influence can be approximated by the expected wavefront error. Using a model based on its root-mean-square  $W_{RMS}$  as in eq. 10 (ALICI et al. 2019; HOLST 2000), the wavefront error is adjusted to match MTF values specified by the optical system manufacturer.

$$MTF_{WFE} = 1 - \left(\frac{W_{RMS}}{0.18}\right)^2 \left[1 - 4 \left(k_n - \frac{1}{2}\right)^2\right] \quad (10)$$

The accumulation of photoelectrons within the physical dimensions of a single detector pixel means that no structures smaller than these dimensions can be recovered. In other words, each

individual pixel presents a square aperture with side length  $w_{px}$ , over which the incoming signal is averaged (ALICI et al. 2019; COTA et al. 2008; WAHBALLAH et al. 2021). While in  $x$  direction this effect is already covered by the projection model, it must be accounted for by an additional term in the model MTF for the  $y$  direction as shown in eq. 11.

$$MTF_{aperture} = \text{sinc}(k_y w_{px}) \quad (11)$$

Note that the extent of a pixel's projection on the terrain in  $y$  direction, as shown in Figure 1, does not represent the physical dimension  $w_{px}$  of the corresponding detector element. Rather, it is caused by the relative movement between the sensor and the imaged object during the exposure. The resulting unavoidable smear in flight direction is therefore also accounted for by the projection model.

However, additional smear components may arise from imperfections in the time-delay and integration (TDI) sensor setup and across-scan velocity components: A mismatch  $\Delta v_y$  between the time between sensor readouts and the velocity of the projected image in the focal plane degrades the image in scan direction. If the direction of flight and charge transfer direction are not perfectly parallel but form an angle  $\theta$ , further degradation in across-scan direction occurs. These aspects are simulated as shown in eq. 12 to 13 (WAHBALLAH et al. 2021).

$$\Delta v_y = \frac{p_y}{t_{int}} - v_y \frac{f}{H} \quad (12)$$

$$MTF_{TDI} = \frac{\sin(\pi N_{TDI} w_{px} \tan(\theta) k_x)}{N_{TDI} \sin(\pi w_{px} \tan(\theta) k_x)} \frac{\sin(\pi N_{TDI} |\Delta v_y| t_{int} k_y)}{N_{TDI} \sin(\pi |\Delta v_y| t_{int} k_y)} \quad (13)$$

with the pitch between TDI stages  $p_y$ , the number of TDI stages  $N_{TDI}$ , and integration time  $t_{int}$ , the duration over which charge is accumulated in the sensor before readout.

Due to random movement of the electrons generated in the detector, the image is blurred further as electrons may eventually arrive in a neighboring pixel well, an effect known as charge diffusion. Different models for a corresponding MTF are available (e.g. COTA et al. 2008; DJITE et al. 2012; SHCHERBACK & YADID-PECHT 2004). In this work, the model presented in eq. 14 will be used (ALICI et al. 2019; FIETE 2010; WAHBALLAH et al. 2021)

$$MTF_{diffusion} = \left(1 - \frac{\exp(-\alpha_{abs} l_{dep})}{1 + \alpha_{abs} l_K}\right) \left(1 - \frac{\exp(-\alpha_{abs} l_{dep})}{1 + \alpha_{abs} l_{diff}}\right)^{-1}, \quad l_K = \frac{l_{diff}}{\sqrt{1 + (2\pi l_{diff} k_r)^2}} \quad (14)$$

with absorption coefficient  $\alpha_{abs}$ , diffusion length  $l_{diff}$ , and depletion width  $l_{dep}$ .

In addition to these effects, high-frequency random motion of the satellite system, so-called jitter, affects image quality. The displacements are modelled by a normal distribution with variance  $\sigma_{jitter}^2$ . The corresponding MTF is therefore a Gaussian given by eq. 15 (ALICI et al. 2019; WAHBALLAH et al. 2021).

$$MTF_{jitter} = \exp(-2\pi^2 \sigma_{jitter}^2 k_r^2) \quad (15)$$

It is important to note that the MTF obtained through composition of the individual steps mentioned here is not equivalent to the overall system MTF, as the steps modelled in the spatial domain – such as downsampling to the simulated image size, or linear smear in flight direction – do not explicitly appear in it.

#### 4.4 Noise

Various sources of noise influence the output image. Photon shot noise and dark current noise are Poisson-distributed, while readout noise and noise generated by electronic circuitry are normally distributed. To simulate the different sources of noise present in the imaging system, additive white Gaussian noise is added to each output pixel and each band independently. The Poisson-distributed noise components are also approximated as normally distributed (BÖRNER et al. 2001; COTA et al. 2008).

The noise standard deviations are adjusted to match signal-to-noise ratios (SNRs) specified by the camera manufacturer for each recorded band. The Gaussian noise is added to the simulated image after application of the MTF as the last step of the simulation process.

### 5 Results & Discussion

Exemplary MTF curves are shown in Figure 2. The influence of the averaging due to  $MTF_{aperture}$ , which is only performed in the frequency domain in  $y$  direction, and instead modelled in the spatial domain in  $x$  direction, can clearly be seen. For the production of output images, the MTF was calculated for each band individually based on its central wavelength, and on a grid spanning  $x$  and  $y$  dimensions. Noise levels are given in Table 1. Simulated imagery as well as real satellite data acquired by a comparable system (PLANET 2022) are shown in Figure 3.

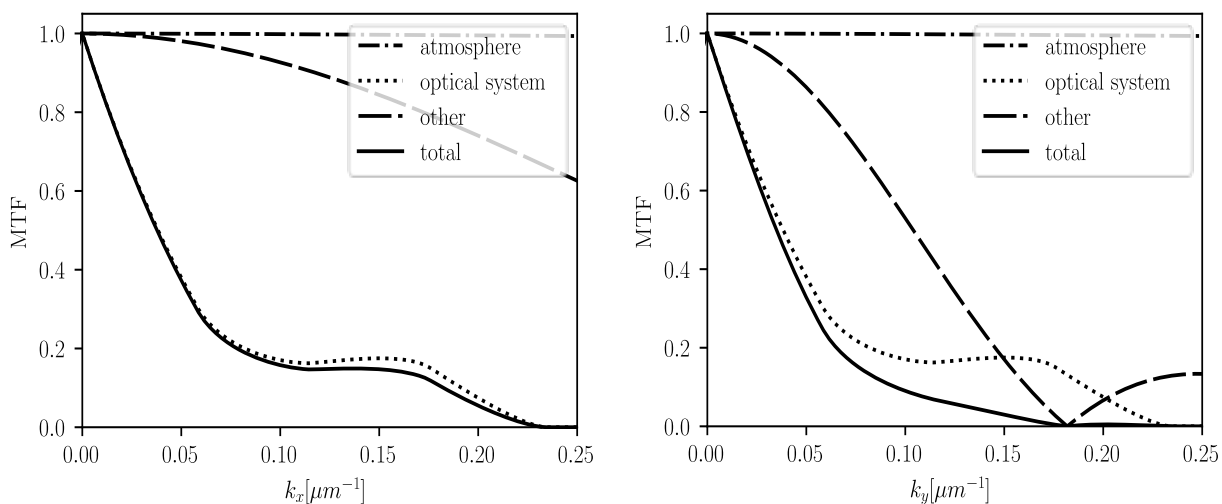


Fig. 2: MTF across (left) and in (right) flight direction

Tab. 1: SNRs and corresponding noise levels

Band	Specified SNR	Mean band value	Noise std. deviation
red	100	102.26	1.02
green	110	110.58	1.01
blue	120	101.21	0.84



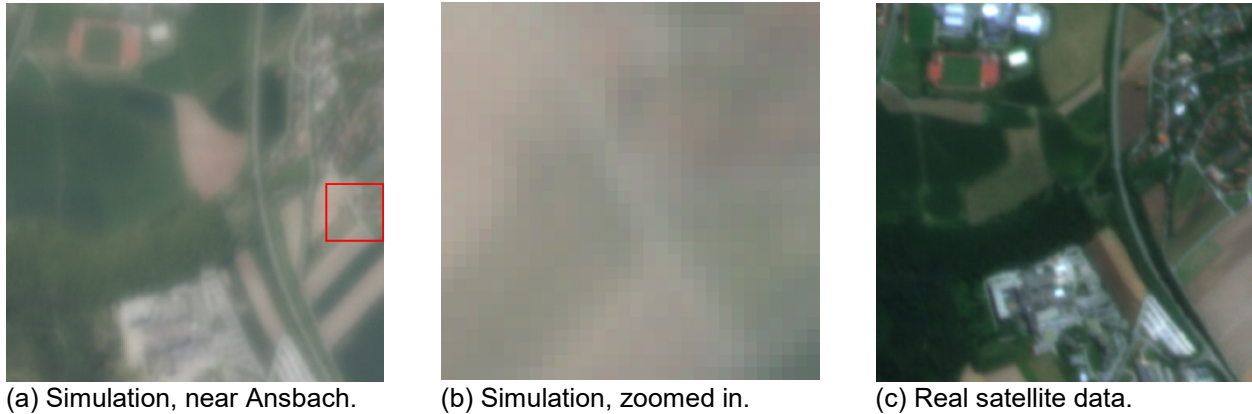


Fig. 3: (a) and (b) show sections of simulated images and (c) actual Planet satellite data to compare

### 5.1 Simulation of Oscillations

While small vibrations at frequencies above the line rate are modelled as jitter, slower deviations from nominal flight conditions may be simulated by applying additional translations and rotations to the orbit data. Sinusoidal oscillations of pitch and roll angles are used to simulate a gyroscopic movement or precession of the satellite. Frequency and amplitude of the oscillations are varied along the satellite's flight to investigate their effects.

Oscillations with peak-to-peak amplitudes  $<0.0005^\circ$  are near imperceptible even in a direct comparison of the simulated images with and without disturbances. Above a certain frequency, the shape of the oscillations becomes less and less pronounced in the simulated images. Instead, the disturbances simply cause additional blurring of the image. For frequencies higher than half of the line rate, aliasing occurs according to the Nyquist theorem. These oscillations should therefore not be modelled as disturbances of the satellite attitude, but rather by taking them into account as jitter. While in the early stages of satellite design the capabilities of the attitude control system cannot be definitively stated yet, a look at other nanosatellite missions may give an idea of the expected accuracy both in terms of tracking error and compensation of unwanted angular velocities. Using reaction wheels, the rotations can be reduced to between  $3.0 \times 10^{-3}$  and  $0.5 \times 10^{-3}$  rad/s (LI et al. 2013a), and tracking errors of  $0.0010^\circ$  to  $0.0012^\circ$  can be achieved (LI et al. 2013b; OLAND & SCHLANBUSCH 2009).

Assuming a similar accuracy for the simulated mission, a positional error of the field of view on the ground surface of a few pixels, or in the order of around 10 m, must be expected. The relatively low angular velocities, however, suggest that a significant degradation of image resolution or visible distortions of the geometry of pictured objects could be avoided. Due to the speculative nature of these considerations, however, the simulations should be repeated once better estimates of the actual attitude control capabilities are available.

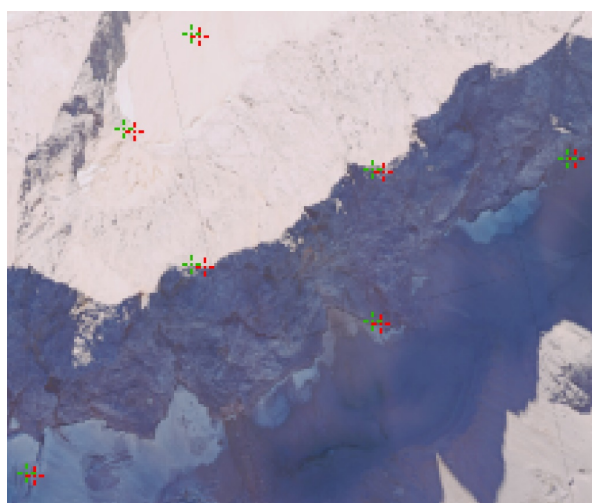
### 5.2 Localization of Ground Control Points

Several GCPs on and below a ridge near the Zugspitze summit were fed into the program to investigate the effect the introduction of a DSM might have on the positional accuracy of the simulation. Each point was given once with its actual elevation (between 2300 m and 2962 m),

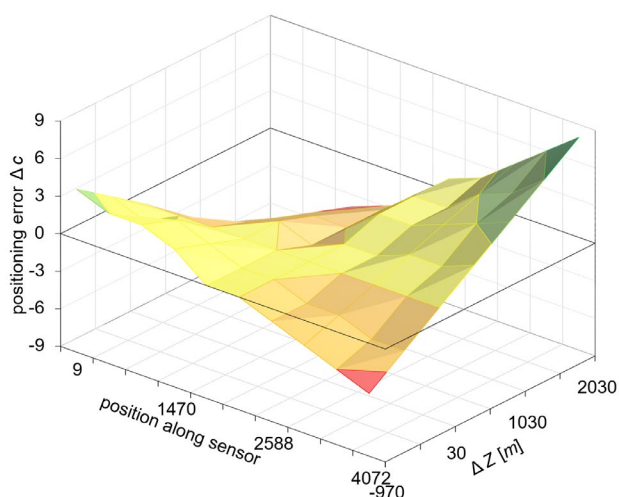
and once assumed to lie in the DOP plane (at 970 m). The resulting positions of exemplary GCPs are displayed in Figure 4a.

To better visualize the contributions to this positioning error, virtual GCPs are generated across the sensor's field of view and at varying elevations. The resulting distribution of the positioning errors  $\Delta c$  in pixels depending on a point's vertical offset from the DOP plane  $\Delta Z$  and its position along the sensor is shown in Figure 4b.

Fitting a bilinear surface to this distribution yields an expected positioning error of 1 m for every 52.30 m of vertical offset at the sensor's edges for the given mission parameters. The bilinear model assumes a perfect nadir view, such that no error can occur in the center of the line sensor. Despite this limitation, it is a useful tool for a rough error estimation under normal flight conditions. These results demonstrate the potential gains in positional accuracy by introduction of a DSM.



(a) Positions of GCPs near Zugspitze summit.



(b) Positioning error distribution.

Fig. 4: (a) Positions of GCPs with correct elevation (green) and GCPs assumed to lie in the DOP plane (red) in a simulated image, and (b) positioning error distribution for fictional GCPs

### 5.3 Radiometric Properties

As discussed before, the radiometric scaling of the input image data is not known. If the prediction of radiometric quantities such as at-sensor radiance is desired, a different set of input data or more details about the original data acquisition are needed. It must be possible to establish a mapping from the original data to either surface reflectance or directly to bottom-of-atmosphere radiance to simulate images with some degree of radiometric accuracy.

Additionally, detailed knowledge of the simulated sensor characteristics, such as the magnitude of the individual components of total sensor noise, or possible non-linearity of the detector response would allow an even more exact simulation.

If the intensity of the output image can be quantified in units of radiance, it is also possible to predict the overall level of noise more accurately. Assuming photon shot noise is the dominating source of noise, the SNR is proportional to the square root of the number of generated

photoelectrons. This relation would allow to scale the noise variance according to the actual at-sensor radiance, instead of assuming a constant SNR for each band.

Without having a dataset as specified above, a quantified comparison between our simulated images and Planet images is not meaningful. Further, the Planet data is preprocessed and not raw satellite imagery. When evaluating the results via qualitative visual inspection our results, as shown in Figure 3, appear foggier with less contrast. On a geometric basis, the images are more similar.

## 6 Conclusion & Outlook

Our work proposes a simulator design adjusted specifically to line scanner imagery acquired by low Earth orbit satellites. Two main areas with potential for future improvements to the simulator have been identified: First, for spectral properties to be represented with any degree of accuracy, input data which allows for the inference of radiometric quantities must be used. Second, the positional accuracy of simulated images could be improved by the introduction of a DSM.

To fully assess the quality of the simulated data, the benefits of post-processing techniques such as MTF compensation should be considered. The MTF model established here may be beneficial to these investigations. Once the satellites are operational, a comparison of the resulting images with those generated by the simulator could allow for an assessment of the methods proposed here and expose further areas for improvement.

## 7 Bibliography

- ALICI, K., OKTEM, F., KARCI, O., YILMAZ, A. & SELIMOGLU, O., 2019: Image chain simulation for earth observation satellites. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **12**(10).
- BÖRNER, A., WIEST, L., KELLER, P., REULKE, R., RICHTER, R., SCHAEPMANN, M. & SCHLÄPFER, D., 2001: SENSOR: a tool for the simulation of hyperspectral remote sensing systems. *ISPRS Journal of Photogrammetry and Remote Sensing*, **55**(5-6), 299-312.
- COTA, S., BELL, J., BOUCHER, R., DUTTON, T., FLORIO, C., FRANZ, G., GRYCEWICZ, T., KALMAN, L., KELLER, R., LOMHEIM, T., PAULSON, D. & WILLKINSON, T., 2008: PICASSO: An End-to-End Image Simulation Tool for Space and Airborne Imaging Systems. *Remote Sensing System Engineering*, Ardanuy, P. & Puschell, J. (Hrsg.), SPIE, Bellingham.
- CHEN, X., LI, T., XUE, B., ZHANG, X., CHEN, G. & NI, G., 2009: Satellite high resolution imaging simulation in space field. *International Symposium on Photoelectronic Detection and Imaging 2009: Advances in Imaging Detectors and Applications*, Zhang, K., Wang, X., Zhang, G. & Ai, K. (Hrsg.), SPIE, Bellingham, 162-171.
- CRUSAN, J. & GALICA, C., 2019: NASA's CubeSat Launch Initiative: Enabling Broad Access to Space. *Acta Astronautica*, **157**, 51-60.
- DJITE, I., ESTRIBEAU, M., MAGNAN, P., ROLLAND, G., PETIT, S. & SAINT-PE, O., 2012: Theoretical models of modulation transfer function, quantum efficiency, and crosstalk for CCD and CMOS image sensors. *IEEE Transactions on Electron Devices*, **59**(3), 729-737.

- FIETE, R., 2010: Modeling the imaging chain of digital cameras. Tutorial texts in optical engineering, **92**, SPIE, Bellingham.
- HOLST, G., 2000: Electro-optical Imaging System Performance. 2<sup>nd</sup> edition. SPIE, Bellingham.
- LI, J., POST, M. & LEE, R., 2013a: Real-time Nonlinear Attitude Control System for Nanosatellite Applications. Journal of Guidance, Control, and Dynamics, **36**(6), 1661-1671.
- LI, J., POST, M., WRIGHT, T. & LEE, R., 2013b: Design of Attitude Control Systems for CubeSat-Class Nanosatellite. Journal of Control Science and Engineering, **2013**(5).
- OLAND, E. & SCHLANBUSCH, R., 2009: Reaction Wheel Design for CubeSats. 4<sup>th</sup> International Conference on Recent Advances in Space Technologies, Kurnaz, S., Ince, F., Onbasioglu, S. & Basturk, S. (Hrsg.), 778-783.
- O'NEILL, E., 1956: Transfer Function for an Annular Aperture. Journal of the Optical Society of America, **46**(4), 285-288.
- PLANET, 2022: Planet Application Program Interface: In Space for Life on Earth. <https://api.planet.com>, letzter Zugriff 10.09.2022.
- SADOT, D. & KOPEIKA, N., 1994: Effects of absorption on image quality through a particulate medium. Applied Optics, **33**(30), 7107-7111.
- SHCHERBACK, I. & YADID-PECHT, O., 2004: CMOS APS MTF Modeling. CMOS Imagers: From Phototransduction to Image Processing, Yadid-Pecht, O. & Etienne-Cummings, R. (Hrsg.), Springer, Dordrecht, 53-74.
- WAHBALLAH, W., LANEVE, G., EL-TOHAMY, F. & BAZAN, T., 2021: Modeling and Simulation of a Remote Sensing Satellite Camera. Sensors, Systems, and Next-Generation Satellites, **25**.

# HEC-RAS-Simulation einer Glacial Lake Outburst Flood zur Ermittlung von möglichen Auswirkungen am Beispiel des Chamlang North Tsho (Lake 464) im Hongu Valley, Nepal

NIELS DEDRING<sup>1</sup>, VALERIE GRAW<sup>1</sup> & ANDREAS RIENOW<sup>1</sup>

*Zusammenfassung: Eine „Glacial Lake Outburst Flood“ (GLOF) beschreibt das rapide Abfließen eines Gletschersees, beispielsweise durch Einsturz und Erosion der ehemaligen Endmoräne, welche den See aufstaut. GLOFs werden in der Regel durch Lawinen und Fels-/Steinschlag oder Kombination verschiedener Ereignisse ausgelöst. Dabei ereignen sich über ein Drittel aller weltweiten GLOFs in der Hindukusch-Himalaya Region. Ziel dieser Arbeit ist es, das GLOF-Risiko und die etwaigen Auswirkungen des Chamlang North Tsho (Lake 464), welcher sich ca. 23 km südlich des Mount Everest befindet, zu analysieren. Hierzu wurden unter anderem die aktuelle Entwicklung des Sees auf Grundlage von Sentinel 1 Aufnahmen betrachtet sowie verschiedene GLOF-Szenarien mittels der Software HEC-RAS entwickelt.*

## 1 Einleitung

Eine sogenannte „Glacial Lake Outburst Flood“ (GLOF) beschreibt den plötzlichen Wasseraustritt eines Gletschersees unabhängig von der Ursache oder dem Auslösemechanismus. Dabei können Abflüsse von mehreren zehntausend Kubikmetern pro Sekunde erreicht und wiederum durch das hohe Erosions- und Transportpotenzial Murgänge bzw. Schlammlawinen mit einem Volumen von über mehreren Millionen Kubikmetern ausgelöst werden. Hauptursache für den Dambruch dieser Seen sind in der Regel gravitative Massenbewegungen, vorrangig Lawinen, aber auch Fels- und Steinstürze, Hangrutschungen, Extremwetterereignisse, Erdbeben oder vorausgehende GLOFs, welche rasch in den Gletschersee stürzen, eine tsunamiähnliche Flutwelle initiieren und durch Überschwappen („Overtopping“) den Damm/Moräne durch Erosion zum Einsturz bringen (ROUNCE et al. 2017). Die hierdurch ausgelöste Flutwelle kann Distanzen von weit über 100 km zurücklegen und stellt eine große sozio-ökonomische Gefahr für die flussabwärts lebende Bevölkerung, die angrenzenden Dörfer und Infrastruktur dar (HARRISON et al. 2018; MASKEY et al. 2020). Nach HARRISON et al. (2018) haben sich mit mehr als einem Drittel aller GLOFs (55 von 160) die meisten in der Hindukush-Himalaya Region ereignet. Ein vergleichsweise unerforschter Gletschersee mit einem wahrscheinlich hohem Risikopotential für eine zukünftige GLOF befindet sich ca. 23 km südlich des Mount Everest im Hongu Valley und ist als „Chamlang North Tsho“ oder „Lake 464“ bekannt (BYERS et al. 2013). Ziel ist es, die aktuelle Entwicklung durch Radarfernerkundung darzustellen und anschließend mit der Wasserhydraulik-Modellierungssoftware HEC-RAS mögliche GLOF-Szenarien und ihre Auswirkungen zu entwickeln. Der Chamlang North Tsho liegt bei 27.7838776° NB und 86.9562533° ÖL in einer Höhe von 5185 m über N. N. im Hongu Valley der Mahakulung Gemeinde und ist damit dem Distrikt Solukhumbu der Provinz Nr. 1, im Osten Nepals zugehörig. Einer der hier möglichen GLOF-Auslöser stellen

---

<sup>1</sup> Ruhr-Universität Bochum, Geographisches Institut, Universitätsstraße 150, D-44801 Bochum, E-Mail: niels.dedring@ruhr-uni-bochum.de

Eisfelder und Überhänge an der Nordwand des Chamlang dar, welche in den See stürzen könnten. Den jedoch mit Abstand häufigsten Auslöser für GLOFs im Himalaya und damit auch für den Chamlang North Tsho bilden Lawinenabgänge (WANG et al. 2011; MOHANTY & MAITI 2021). Nach LALA et al. (2018) weisen Hangneigungen von 45° bis 60° eine erhöhte Lawinengefährdung auf, was übertragen auf den Chamlang North Tsho eine gefährdete Fläche von insgesamt ca. 2,69 km<sup>2</sup> ergibt.

## 2 Methodologie

Für die nachfolgende Analyse des Chamlang North Tsho hinsichtlich seiner aktuellen Größe und Entwicklung in den letzten Jahren sowie der GLOF-Simulation wurden ausschließlich kostenfreie und frei verfügbare Daten verwendet. Die Analyse ist dabei unterteilt in zwei Teilanalysen. Erstens der Ermittlung der aktuellen Ausdehnung des Sees und seiner Entwicklung der letzten Jahre (Kapitel 2.1) und zweitens einer GLOF-Modellierung und Simulation des Chamlang North Tsho (Kapitel 2.2). Die beiden Kapitel sind aufeinander aufbauend und die Ergebnisse aus Kapitel 2.1 werden für die Simulation in Kapitel 2.2 verwendet.

### 2.1 Aktuelle Entwicklung des Chamlang North Tsho

#### 2.1.1 Daten

Bei den verwendeten Daten handelt es sich um High Resolution Level-1 GRD Produkte des Sentinel-1A Satelliten aus den Jahren 2014 bis 2021. Für die Jahre 2015 bis 2021 stammen die Aufnahmen aus Anfang bis Ende September. Eine Ausnahme bildet das Jahr 2014, da Sentinel-1A in diesem Jahr gestartet ist und die ersten Bilder des Untersuchungsgebietes erst im Oktober aufgenommen wurden. Zwischen Mitte November und Mitte Juni ist der See vollständig oder zumindest partiell von Eis bedeckt, wodurch bei dem gewählten Zeitraum (September und Oktober) eine Eisbedeckung ausgeschlossen werden kann (SENTINEL HUB 2021a; 2021b).

#### 2.1.2 Methodik

Zur Vorverarbeitung der Sentinel-1A Daten wurde die Open Source Softwareanwendung SNAP verwendet. Die Abbildung 1 stellt die einzelnen Schritte schematisch dar. Das Vorgehen orientiert sich dabei an die empfohlene Praxis zur Hochwasserkartierung von UN-SPIDER (o.J.) und FILIPPONI (2019).



Abb. 1: Workflow zur Vorverarbeitung der Sentinel-1A Daten

## 2.2 Glacial Lake Outburst Flood Modellierung mit HEC-RAS

### 2.2.1 Daten

Kernkomponente für eine einfache GLOF-Simulation besteht unter anderem aus einem digitalen Geländemodell (DGM). Das hier verwendete DGM basiert auf radiometrisch geländekorrigierten Daten von ALOS PALSAR und wird von der Alaska Satellite Facility mit einer Auflösung von  $12,5\text{ m} \times 12,5\text{ m}$  zur Verfügung gestellt (ASF 0.J.). Zur Darstellung der Landbedeckung wurde das ESA WorldCover 2020 verwendet. Es bietet eine globale Abdeckung mit einer Auflösung von  $10\text{ m} \times 10\text{ m}$  (ESA 2020). Diesem wurden Rauigkeitskoeffizienten basierend auf der MS-GAP und NLCD-Klassifizierung zugeordnet, da derzeit noch keine dedizierten Koeffizienten existieren (BUNYA et al. 2010). Ferner werden Daten zur aktuellen Seeoberfläche, Tiefe und Volumen des Sees benötigt bzw. hier aus den Ergebnissen aus Kapitel 2.1 numerisch ermittelt. Als Auslöser dienen drei verschiedene Lawinenszenarien, mit einer kleinen ( $5 \times 10^4\text{ m}^3$ ), einer mittleren ( $9 \times 10^5\text{ m}^3$ ) und einer großen ( $6,6 \times 10^6\text{ m}^3$ ) Lawine. Selbige Größen wurden von LALA et al. (2018) für eine GLOF Modellierung des nahegelegenen Imja Lake verwendet. Zur Ermittlung potenzieller Auswirkungen und Schäden an Gebäuden, Brücken sowie Straßen und Wegen, werden anschließend OpenStreetMap-Daten mit den Ergebnissen der GLOF-Simulation verschnitten.

### 2.2.2 Methodik

Das schematische Vorgehen der HEC-RAS Analyse wird in Abbildung 2 vereinfacht dargestellt.

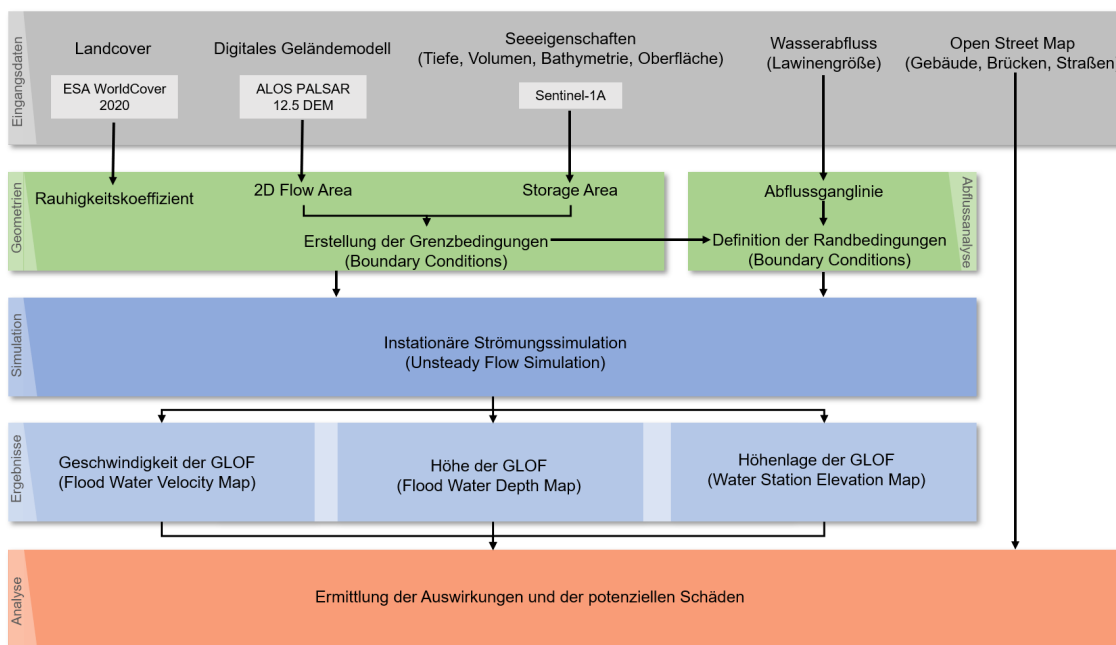


Abb. 2: Schematischer Ablauf der Analyse mit HEC-RAS

Die mittlere Tiefe wurde nach FUJITA et al. (2013) auf 53,45m ermittelt und durch den Vergleich mit dem benachbarten Chamlang South Tsho auf eine maximale Tiefe von 90 m geschätzt. Das Volumen wurde nach dem empirischen Modell von YAO et al. (2012) auf  $44336925,34\text{ m}^3$

bestimmt. Auf Basis der ermittelten Parameter (Tiefe, Volumen) wurde eine exponentielle Volumen-/Höhenkurve erstellt und der Seegeometrie (Storage Area) zugeordnet. Das Zellennetz der 2D Flow Area wurde mit einer Auflösung von  $20 \text{ m} \times 20 \text{ m}$  sowie einer detaillierteren Auflösung von  $10 \text{ m} \times 10 \text{ m}$  im unmittelbaren Bereich des Flusslaufes generiert. Damit werden insgesamt 309754 einzelnen Zellen zur Berechnung der Simulation verwendet. Für die Berechnung eines Dammbrechens des Chamlang North Tsho wurde die Methode nach FROEHLICH (2008) verwendet. Hierbei wurde von einem „overtopping“ der Endmoräne und einer sinusartigen Erosion ausgegangen. Der hier berechnete Dammbbruch besitzt eine errechnete finale Dammbbruchbreite von 83 m, eine maximale Tiefe von 20 m, eine seitliche Neigung von 1, einen Bruch-Wehrkoeffizienten von 1,44 und eine gesamte Einsturzzeit von 1,41 Stunden.

### 3 Ergebnisse und Diskussion

Die Ergebnisse aus Kapitel 2.1 sind in Abb. 3 dargestellt. Aufgrund von diversen Bildartefakten musste die Ermittlung des aktuellen Umfangs des Sees händisch erfolgen und konnte nicht automatisiert durchgeführt werden. Dabei ist ein geringer, aber positiver Anstieg in der Fläche zu erkennen. Trotz der damit verbundenen Subjektivität sowie der Bildartefakte hat sich die Radartechnik für diesen Untersuchungsraum als zuverlässig erwiesen. Dabei lagen hier essenzielle Vorteile gegenüber der optischen Fernerkundung in der Unabhängigkeit der Witterungsverhältnisse, insbesondere zur Monsunzeit, sowie in der Unabhängigkeit des Schattenwurfs des Chamlangs.

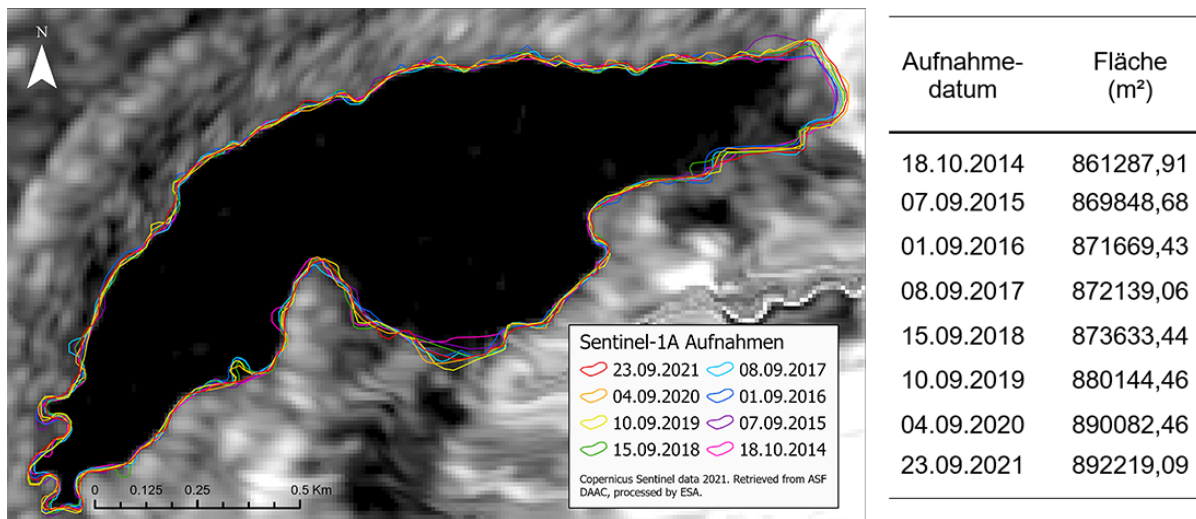


Abb. 3: Fläche des Chamlang North Tsho von 2014 bis 2021

Die Berechnung der GLOF-Simulation dauerte in Abhängigkeit der drei Lawinen-Szenarien, 7,3 Stunden, 13,65 Stunden bzw. 35,73 Stunden. Die endgültige geometrische Ausbreitung der Überflutungsszenarien sowie das ermittelte Schadensausmaß auf Basis von OSM-Daten sind in Abbildung 4 sowie Tabelle 1 aufgeführt. Eine Validierung der Ergebnisse ist unter anderem auf Grundlage mangelnder Datenlage sowie im Rahmen dieser Arbeit nicht möglich gewesen. Die



einzigsten Studien, welche sich mit dem Chamlang North Tsho auseinandergesetzt haben, gehen auf BYERS et al. (2013), ROUNCE et al. (2016) und ROUNCE et al. (2017) zurück. Beim Vergleich der

Tab. 1: Räumliche Ausdehnung und potenzielle Schäden der drei GLOF Simulationen.

Lawinengröße (m <sup>3</sup> )	Entfernung zum See (m)	Fläche (km <sup>2</sup> )	Gebäude	Brücken	Straßen/Wege (km)
5 x 10 <sup>4</sup>	22,787	1,894	0	0	0,123
9 x 10 <sup>5</sup>	109,753	12,312	71	22	7,051
6,6 x 10 <sup>6</sup>	147,256	35,729	235	30	28,752

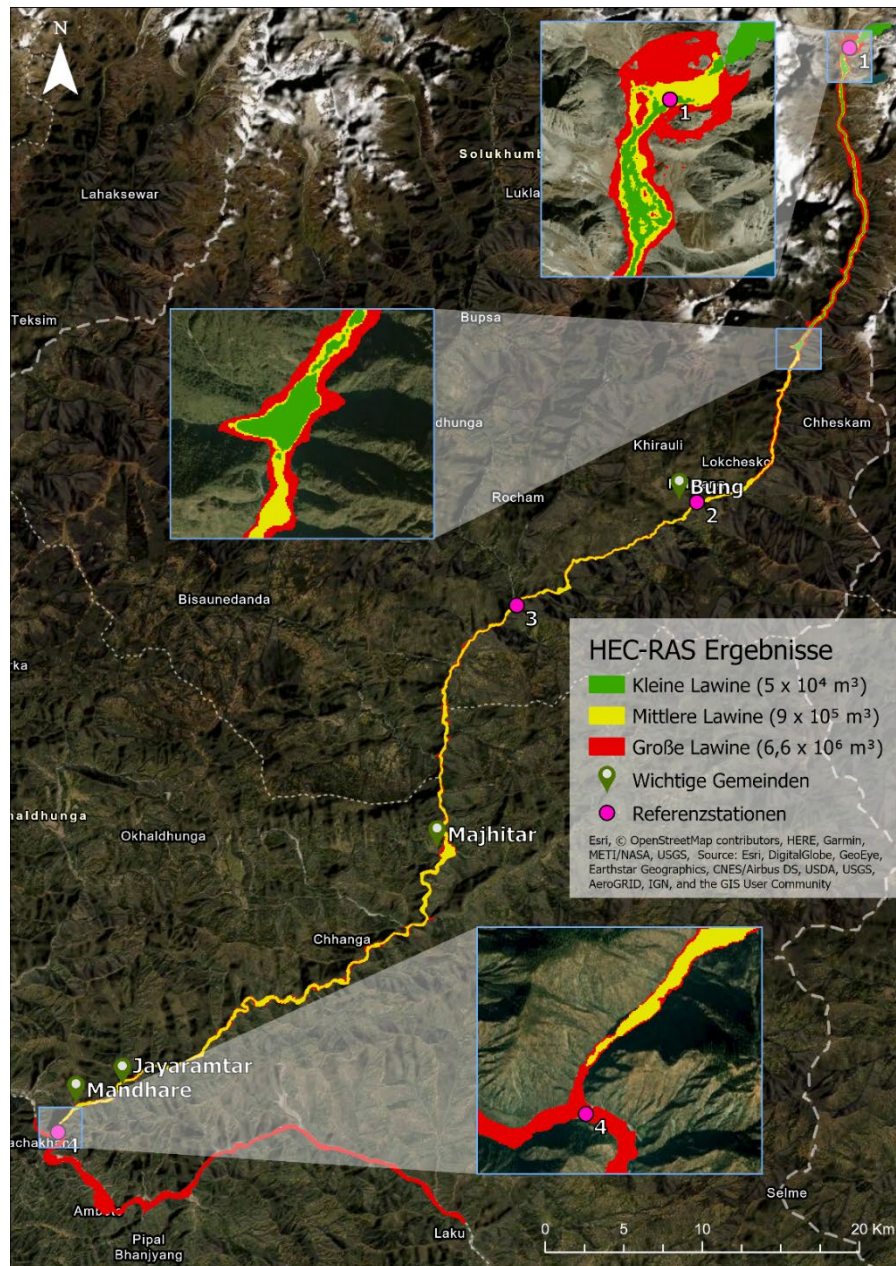


Abb. 4: Die Ergebnisse der GLOF-Simulationen.

Ergebnisse mit diesen Studien weist die Simulation unter dem mittleren Lawinen-Szenario die größte Annäherung an dessen Ergebnisse vor. Dabei kommt es in Teilaspekten der Simulation sowohl zur Über- als auch zu Unterschätzungen. Die Ähnlichkeit des mittleren-Szenarios mit anderen Studien zeigt jedoch auch, dass die Modellierung des kleinen Lawinen-Szenarios wahrscheinlich deutlich unterschätzt bzw. das große Lawinen-Szenario deutlich überschätzt. Unter Annahme der Ergebnisse konnte in Verbindung mit OSM-Daten ein grobes potenzielles Schadensbild ermittelt werden. Dieses kann zwar nicht vollumfänglich die gesamte Infrastruktur abbilden, aber dennoch erste Aussagen über die Vulnerabilität in den angrenzenden Dörfern zulassen. So weisen beispielsweise die Gemeinden „Bung“ und „Majhitar“, durch die Nähe zum See bzw. durch landwirtschaftliche Nutzung in unmittelbarer Flussnähe, eine hohe Vulnerabilität gegenüber einer GLOF auf. Zur Katastrophenvorsorge wäre eine Möglichkeit die Errichtung eines Frühwarnsystems in Verbindung mit einer Überwachungsstation am Chamlang North Tsho. Eine weitere Möglichkeit zur Risikominimierung stellt das Absenken des Wasserspiegels durch kontrolliertes Abtragen der Endmoräne dar. Dieses Vorgehen hat sich bereits am Thorthormi See im Buthan bewährt (WORLD WIDE FUND OF NATURE 2009). Für diese Studie mussten zahlreiche Annahmen, insbesondere durch mangelnde In-Situ Daten, getroffen werden, welche schlussendlich die Aussagekraft der Simulation limitiert. Es fehlt beispielsweise eine exakte Vermessung der Bathymetrie des Sees zur Volumenermittlung sowie der Zusammensetzung und Geometrie der Endmoräne. Des Weiteren fehlen genaue Rauigkeitskoeffizienten, sowie ein höher aufgelöstes DGM für das Untersuchungsgebiet. Mit ebendiesen Daten wäre es möglich, eine exaktere und komplexere Modellierung durchzuführen, um schlussendlich validere Ergebnisse zu erhalten.

#### **4 Fazit und Ausblick**

In dieser Arbeit konnte ein vereinfachter Weg aufgezeigt werden, eine GLOF-Simulation unter Verwendung von ausschließlich frei verfügbaren Daten und Softwareanwendungen (HEC-RAS) zu erstellen und erste Aussagen über dessen Verlauf und mögliche Auswirkungen abzuleiten. Diese Ergebnisse sind bereits ausreichend, um erste Schritte für Präventions- und Anpassungsmaßnahmen vor Ort einzuleiten. Dennoch mussten aufgrund mangelnder Datenverfügbarkeit diverse Annahmen getroffen werden, was wiederum zu Einschränkungen und Limitationen in den Ergebnissen führte. Um präzisere bzw. validere Aussagen treffen zu können, ist eine Expedition zum Chamlang North Tsho zur näheren Feldforschung unerlässlich. Im Fokus sollte eine genaue Vermessung der Bathymetrie des Sees sowie die Untersuchung des genauen Aufbaus der Endmoräne stehen. Darüber hinaus empfiehlt sich eine drohnengestützte Kartierung des Sees inklusive der näheren Umgebung, um anschließend ein hochauflösendes DGM und ein 3D-Modell zu generieren.

## 5 Literaturverzeichnis

- ALASKA SATELLITE FACILITY (ASF), o. J.: ALOS PALSAR – Radiometric Terrain Correction. <https://asf.alaska.edu/data-sets/derived-data-sets/alos-palsar-rtc/alos-palsar-radio-metric-terrain-correction/>, letzter Zugriff 01.02.2022.
- BUNYA, S., DIETRICH, J., WESTERINK, J., EBERSOLE, B., SMITH, J., ATKINSON, J., JENSEN, R., RESIO, D., LUETTICH, R., DAWSON, C., CARDONE, V., COX, A., POWELL, M., WESTERINK, H. & ROBERTS, H., 2010: A High-Resolution Coupled Riverine Flow, Tide, Wind, Wind Wave, and Storm Surge Model for Southern Louisiana and Mississippi. Part I: Model Development and Validation. *Monthly Weather Review*, **138**(2), 345-377, <https://doi.org/10.1175/2009MWR2906.1>.
- BYERS, A., MCKINNEY, D., SOMOS-VALENZUELA, M., WATANABE, T. & LAMSAL, D., 2013: Glacial lakes of the Hinku and Hongu valleys, Makalu Barun National Park and Buffer Zone, Nepal. *Natural Hazards*, **69**(1), 115-139, <https://doi.org/10.1007/s11069-013-0689-8>.
- EUROPEAN SPACE AGENCY (ESA), 2020: WorldCover Product User Manual. [https://worldcover2020.esa.int/data/docs/WorldCover\\_PUM\\_V1.1.pdf](https://worldcover2020.esa.int/data/docs/WorldCover_PUM_V1.1.pdf), letzter Zugriff am 01.02.2022.
- FILIPPONI, F., 2019: Sentinel-1 GRD Preprocessing Workflow. *MDPI Proceedings*, **18**(11), 1-4, <https://doi.org/10.3390/ECRS-3-06201>.
- FROEHLICH, D. 2008: Embankment Dam Breach Parameters and Their Uncertainties. *Journal of Hydraulic Engineering*, **134**(12), 1708-1721, [https://doi.org/10.1061/\(ASCE\)0733-9429\(2008\)134:12\(1708\)](https://doi.org/10.1061/(ASCE)0733-9429(2008)134:12(1708)).
- FUJITA, K., SAKAI, A., TAKENAKA, S., NUIMURA, T., SURAZAKOV, A., SAWAGAKI, T. & YAMANOKUCHI, T., 2013: Potential flood volume of Himalayan glacial lakes. *Natural Hazards and Earth System Science*, **13**(7), 1827-1839, <https://doi.org/10.5194/nhess-13-1827-2013>.
- HARRISON, S., KARGEL, J., HUGGEL, C., REYNOLDS, J., SHUGAR, D., BETTS, R., EMMER, A., GLASSER, N., HARITASHYA, U., KLIMEŠ, J., REINHARDT, L., SCHAUB, Y., WILTSHIRE, A., REGMI, D. & VILÍMEK, V., 2018: Climate change and the global pattern of moraine-dammed glacial lake outburst floods. *The Cryosphere*, **12**(4), 1195-1209, <https://doi.org/10.5194/tc-12-1195-2018>.
- LALA, J., ROUNCE, D. & MCKINNEY, D., 2018: Modeling the glacial lake outburst flood process chain in the Nepal Himalaya: reassessing Imja Tsho's hazard. *Hydrology and Earth System Sciences* **22** (7), 3721-3737, <https://doi.org/10.5194/hess-22-3721-2018>.
- MASKEY, S., KAYASTHA, R. & KAYASTHA, R., 2020: Glacial Lakes Outburst Floods (GLOFs) modelling of Thulagi and Lower Barun Glacial Lakes of Nepalese Himalaya. *Progress in Disaster Science*, **7**(100106), 1-13, <https://doi.org/10.1016/j.pdisas.2020.100106>.
- MONHANTY, L. & MAITI, S., 2021: Probability of glacial lake outburst flooding in the Himalaya. *Resources, Environment and Sustainability*, **5**(100031), 1-13, <https://doi.org/10.1016/j.resenv.2021.100031>.
- ROUNCE, D., MCKINNEY, D., LALA, J., BYERS, A. & WATSON, C., 2016: A new remote hazard and risk assessment framework for glacial lakes in the Nepal Himalaya. *Hydrology and Earth System Sciences*, **20**(9), 3455-3475, <https://doi.org/10.5194/hess-20-3455-2016>.
- ROUNCE, D., WATSON, C. & MCKINNEY, D., 2017: Identification of Hazard and Risk for Glacial Lakes in the Nepal Himalaya Using Satellite Imagery from 2000–2015. *Remote Sensing*, **9**(7), 654-673, <https://doi.org/10.3390/rs9070654>.

- SENTINEL HUB, 2021a: Sentinel-2 L2A. True Color. 2021-06-11. [https://apps.sentinel-hub.com/eo-browser/?zoom=15&lat=27.78658&lng=86.95868&themeId=DE-FAULT-THEME&visualizationUrl=https%3A%2F%2Fservices.sentinel-hub.com%2Fogc%2Fwms%2Fbd86bcc0-f318-402b-a145-015f85b9427e&datasetId=S2L2A&fromTime=2021-06-11T00%3A00%3A00.000Z&toTime=2021-06-11T23%3A59%3A59.999Z&layerId=1\\_TRUE\\_COLOR](https://apps.sentinel-hub.com/eo-browser/?zoom=15&lat=27.78658&lng=86.95868&themeId=DE-FAULT-THEME&visualizationUrl=https%3A%2F%2Fservices.sentinel-hub.com%2Fogc%2Fwms%2Fbd86bcc0-f318-402b-a145-015f85b9427e&datasetId=S2L2A&fromTime=2021-06-11T00%3A00%3A00.000Z&toTime=2021-06-11T23%3A59%3A59.999Z&layerId=1_TRUE_COLOR), letzter Zugriff 21.01.20220.
- SENTINEL HUB, 2021b: Sentinel-2 L2A. True Color. 2021-11-28. [https://apps.sentinel-hub.com/eo-browser/?zoom=15&lat=27.78656&lng=86.95868&themeId=DEFAULT-THEME&visualizationUrl=https%3A%2F%2Fservices.sentinel-hub.com%2Fogc%2Fwms%2Fbd86bcc0-f318-402b-a145-015f85b9427e&datasetId=S2L2A&fromTime=2021-11-28T00%3A00%3A00.000Z&toTime=2021-11-28T23%3A59%3A59.999Z&layerId=1\\_TRUE\\_COLOR](https://apps.sentinel-hub.com/eo-browser/?zoom=15&lat=27.78656&lng=86.95868&themeId=DEFAULT-THEME&visualizationUrl=https%3A%2F%2Fservices.sentinel-hub.com%2Fogc%2Fwms%2Fbd86bcc0-f318-402b-a145-015f85b9427e&datasetId=S2L2A&fromTime=2021-11-28T00%3A00%3A00.000Z&toTime=2021-11-28T23%3A59%3A59.999Z&layerId=1_TRUE_COLOR), letzter Zugriff 21.01.2022.
- UN-SPIDER, o.J.: Step by Step: Recommended Practice Flood Mapping. <https://www.unspider.org/advisory-support/recommended-practices/recommended-practice-flood-mapping/step-by-step>, letzter Zugriff 26.11.2022.
- WANG, W., WEICAI, Y., TANDONG, GAO, Y., XIAOXIN, Y. & KATTEL, D., 2011: A First-order Method to Identify Potentially Dangerous Glacial Lakes in a Region of the Southeastern Tibetan Plateau. *Mountain Research and Development*, **31**(2), 122-130, <https://doi.org/10.1659/MRD-JOURNAL-D-10-00059.1>.
- WORLD WIDE FUND FOR NATURE, 2009: The Cost of Climate Change. The Story of Thorthormi Lake in Bhutan. [https://wwf.fi/app/uploads/h/5/7/5rqxrm5dl5ejasicpp5m4zc/the\\_cost\\_of\\_climate\\_change.pdf](https://wwf.fi/app/uploads/h/5/7/5rqxrm5dl5ejasicpp5m4zc/the_cost_of_climate_change.pdf), letzter Zugriff 27.11.2022.
- YAO, X., LIU, S., SUN, M., WEI, J. & GUO, W., 2012: Volume calculation and analysis of the changes in moraine-dammed lakes in the north Himalaya: a case study of Longbasaba lake. *Journal of Glaciology*, **58**(210), 753-760, <https://doi.org/10.3189/2012JoG11J048>.

# Spot the Difference: Learned DSM Updating

BINGXIN KE<sup>1</sup>, CORINNE STUCKER<sup>1</sup> & KONRAD SCHINDLER<sup>1</sup>

*Abstract: In a rapidly changing world, keeping 3D city models up to date is a crucial task for many applications. Very high-resolution (VHR) optical satellite sensors, stereo matching algorithms, and learning-based refinement enable the reconstruction of high-quality city models from space, but so far did not account for the evolution of urban scenes. In order to rapidly update an existing city model with a limited set of newly collected satellite images, we introduce a DSM updating method based on neural implicit occupancy fields. We demonstrate that our method is able to effectively combine the old DSM and the new data. Consequently, changed areas can be updated to match new image observations, thus reducing the associated height errors (MAE) by  $\approx 50\%$  compared to the obsolete DSM; while the reconstruction quality in unchanged areas is ensured by the old DSM that is based on more imagery and thus a higher 3D point density.*

## 1 Introduction

Reconstructing and maintaining an up-to-date city-scale scene is a fundamental and important task of photogrammetry and computer vision. The reconstructed high-resolution digital surface model (DSM) serves as a basis for various downstream applications, including topographic mapping, "digital twins", environmental simulations, planning, etc. Nowadays, very high-resolution satellite sensors enable the acquisition of optical remote sensing images with fine-grained details at almost any location on the Earth from different viewpoints in space, within a short time interval. The 3D reconstruction from remote sensing images is possible with the help of tailored stereo matching algorithms (ROTHERMEL et al., 2012; YOUSSEFI et al., 2020).

Due to limitations in terms of image resolution, geometric conditions, and radiometric consistency, the derived point clouds and DSMs tend to be noisy and sometimes incomplete. Recently, learning-based methods have been developed to refine the raw DSM from stereo reconstruction (BITTNER et al. 2018; STUCKER & SCHINDLER, 2022). We build on the *ImpliCity* method (STUCKER et al. 2022), which directly converts raw point clouds into a city model with smooth surfaces, fine-grained shape details, and crisp building edges.

Such existing methods aim to reconstruct DSMs without considering the acquisition time of the input images, and thus without regard for the evolution of the city over time. In actual fact cities do change, mainly due to construction activities. Still, most of the area in an urban scene remains static over a relatively long period, hence an old city model is expected to provide additional data redundancy in unchanged areas, especially in the practical scenario where one aims to keep the model up to date, and therefore update shortly after the change, when only few new satellite images have been collected.

---

<sup>1</sup> ETH Zurich, Photogrammetry and Remote Sensing, Stefano-Franscini-Platz 5, CH-8093 Zurich, E-Mail: [bingke, stuckerc, schindler]@ethz.ch

In this work, we propose a DSM updating method based on satellite images, using a neural implicit surface representation. With the help of a change detection module and properly pretrained encoders and decoder, the proposed method can leverage the existing DSM and newly collected data together to produce high-quality reconstructions of both changed and unchanged areas.

## 2 Methodology

### 2.1 Method Overview

**Problem Formulation.** The old DSM  $D_1$  can be derived from a stack of satellite images  $I_1^{raw}$  that represent the state at time  $T_1$ , with conventional semi-global matching (ROTHERMEL et al., 2012) followed by surface reconstruction (STUCKER et al. 2022), as shown in the left half of Figure 1. Our goal is to update the old DSM state  $D_1$  to a new state  $D_2$ , given only few ( $\geq 2$ ) newly collected satellite images  $I_2^{raw}$  that represent the new state at time  $T_2$ . The updated DSM  $D_2$  is expected to be up to date in the changed areas, but still have the same quality in unchanged areas.

**DSM updating method.** We decompose the problem into two sub-tasks: (1) determining changed areas; (2) fusing data and reconstructing the DSM by making use of all data in the unchanged areas, while ignoring the old DSM in changed areas. As shown in the right half of Figure 1, a new point cloud  $P_2$  representing the new state is derived from the newly collected images  $I_2^{raw}$ . Our method then takes as input the old DSM  $D_1$ , the point  $P_2$ , and ortho-rectified images  $I_1^{D_1}$  and  $I_2^{D_1}$ . The change detector  $g_\phi$  takes one old state image  $I_1^{D_1}$  and one new state image  $I_2^{D_1}$  as input and outputs a binary change mask  $M_{change}$  with pixel-wise change probabilities  $p_i \in \{0,1\}$ , namely,  $g_\phi(I_1^{D_1}, I_2^{D_1}) = M_{change}$ . Having this predicted change mask and other input data (old DSM  $D_1$ , new point cloud  $P_2$ , and new state image pair  $I_2^{D_1}$ ), the DSM updater  $F_\theta$  generates the new DSM  $D_2$ . Thus, our method can be formulated as:

$$F_\theta \left( P_2, I_2^{D_1}, D_1, g_\phi(I_1^{D_1}, I_2^{D_1}) \right) = D_2 \quad (1)$$

The change detector  $g_\phi$  and the DSM updater  $F_\theta$  are parameterized as deep neural networks. Note that due to the ortho-rectification of the imagery,  $D_1$ ,  $M_{change}$ ,  $I_1^{D_1}$ , and  $I_2^{D_1}$  are inherently aligned in the same geographic coordinate system. I.e., the  $(x, y)$ -axes are the East and North directions in the local UTM zone, and the  $z$ -axis is the vertical.

### 2.2 Change Detector

Our change detector module is a Fully Convolutional Siamese architecture, following (DAUDT et al. 2018). Data is cropped into patches for training and inference. To relieve the bias caused by radiometric conditions, the patches are randomly rotated by  $\alpha \in \{0^\circ, 90^\circ, 180^\circ, 270^\circ\}$  and randomly flipped along  $x$  or  $y$  axis during training. We further randomly swap the two input images to avoid asymmetries. Since it is in our context important to cover all changed content, even at the cost of including some unchanged parts, we trade precision for higher recall in post-processing: we perform erosion with radius  $r = 10$  pixels (respectively, 2.5 m) to remove noise, followed by dilation with  $r = 40$  pixels to cover slightly larger and more complete changed areas.

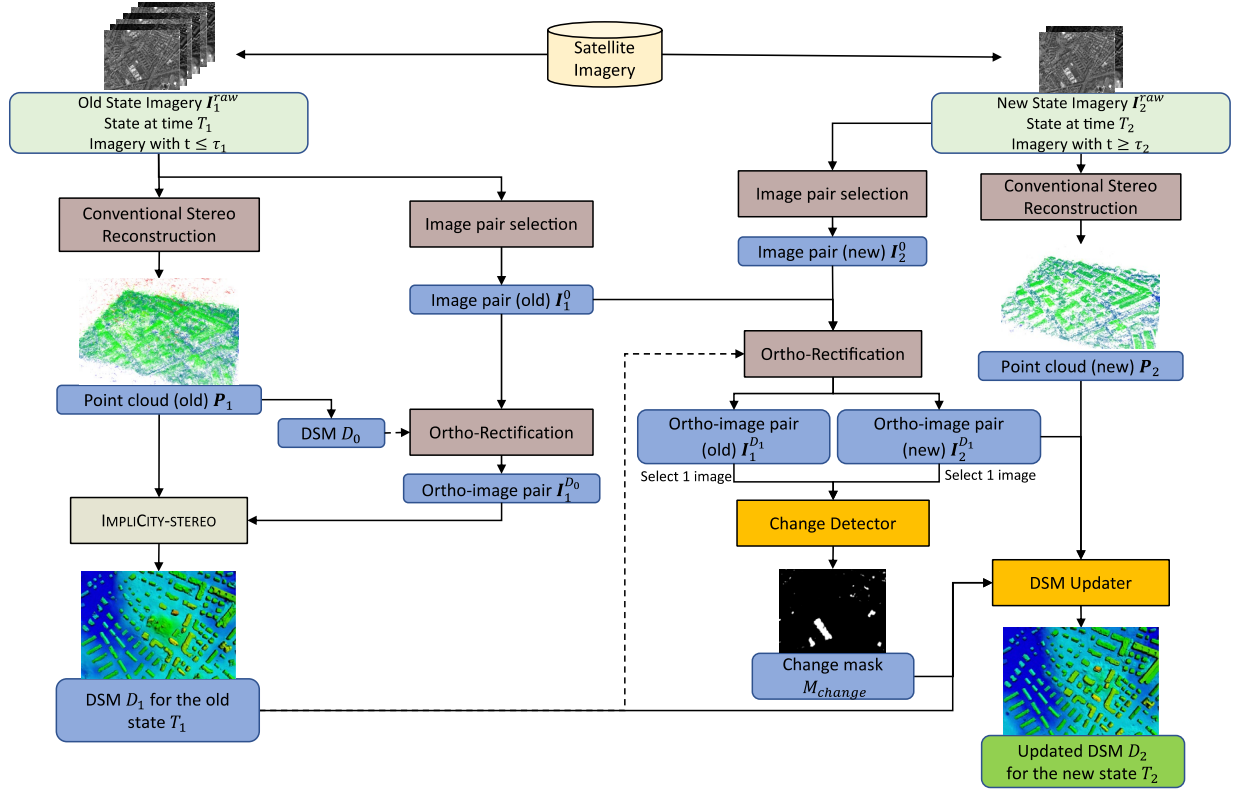


Fig. 1: Method overview

## 2.3 DSM Updater

The DSM updater module is the core of our method. It is adapted from the image-guided, coordinate-based neural representation of *ImpliciCity*. The scene is represented as a convolutional neural occupancy field (PENG et al., 2020), i.e., a function  $f_{\theta}$  that, for any given query point  $x \in R^3$ , returns the corresponding occupancy probability  $\hat{\delta}$ . In our case,  $\hat{\delta} = 0$  for locations above the surface and  $\hat{\delta} = 1$  for points underneath the surface.

**Formulation.** As shown in Figure 2, our network is an encoder-decoder design consisting of two raster encoders (DSM encoder  $f_{DSM}$  and image encoder  $f_{image}$ ), one point cloud encoder  $f_{PC}$ , one feature fusion module  $f_{fusion}$ , and one decoder  $f_{decode}$ . The encoders ( $f_{DSM}$ ,  $f_{image}$ , and  $f_{PC}$ ) convert the input into feature embeddings. For a specific  $(x, y)$  location, the corresponding features and the change probability can be queried by bilinear interpolation from the embeddings. Then the occupancy value is predicted by the decoder. The prediction at a query point  $x$  can be written as:

$$f_{\theta} = f_{decode} \left( x, f_{fusion} \left( f_{DSM}(D_1, x), f_{PC}(P_2, x), M_{change}(x) \right), f_{image}(I_2^{D_1}, x) \right) \rightarrow \hat{\delta} \in [0,1] \quad (2)$$

**Network Components.** The **point cloud encoder** follows the encoder architecture of *ConvONet* (PENG et al., 2020) and convert the input point cloud  $P_2$  into a high-dimensional feature embedding. The **image encoder** is the same as that in *ImpliciCity*. It converts ortho-rectified images into an image feature embedding, to provide additional high-frequency details that would be missed in the sparser point cloud. The **DSM encoder** converts the old DSM into a feature embedding. It shares the same architecture as the image encoder, except for the first layer, which accepts single-

channel input. The **decoder** predicts the occupancy value from the sum of queried features ( $\psi_{fused}$  and  $\psi_{image}$ ) and the coordinate of query point  $\mathbf{x}$ .

**Feature Fusion.** To smartly leverage the input data, we apply a feature-level fusion. Through ablation studies, we conclude that fusing the geometric features (i.e., DSM feature  $\psi_{DSM}$  and point cloud feature  $\psi_{PC}$ ) yields the best result. Note that in changed areas (where  $p = 1$ ), the DSM should be ignored, thus the fusion is disabled and  $\psi_{PC}$  replaces  $\psi_{fusion}$ .

**Training.** During training, query points  $\{\mathbf{x}_i \in R^3\}$  are randomly sampled within the volume of interest, with higher sampling density in the vicinity of building and terrain surfaces. The ground truth occupancy values are assigned according to the relative position of the sample point w.r.t. the ground truth DSM. The network is then trained by minimizing the cross-entropy loss  $\mathcal{L}$  between predicted occupancies  $\hat{o}$  and true occupancies  $o$ :

$$\mathcal{L}(\hat{\mathbf{o}}, \mathbf{o}) = \sum_i (o_i \cdot \log(\hat{o}_i) + (1 - o_i) \cdot \log(1 - \hat{o}_i)) \quad (3)$$

**Pretrained Weights.** Using pretrained network modules yielded excellent reconstruction results in our ablation studies. Thus, in our final model, we use the pretrained weights to initialize the model and fine-tune them to the specific scene with a rather small learning rate.

**DSM Generation.** During inference, the updated DSM can be extracted from the neural occupancy field, as the iso-surface at  $f_\theta = 0.5$ . We do this in a coarse-to-fine search for efficiency. First, query points are coarsely sampled along the  $z$ -axis at every  $(x,y)$  raster cell of the output DSM. Then the  $z$ -range is recursively partitioned around the value  $f_\theta = 0.5$  to find the precise surface height.

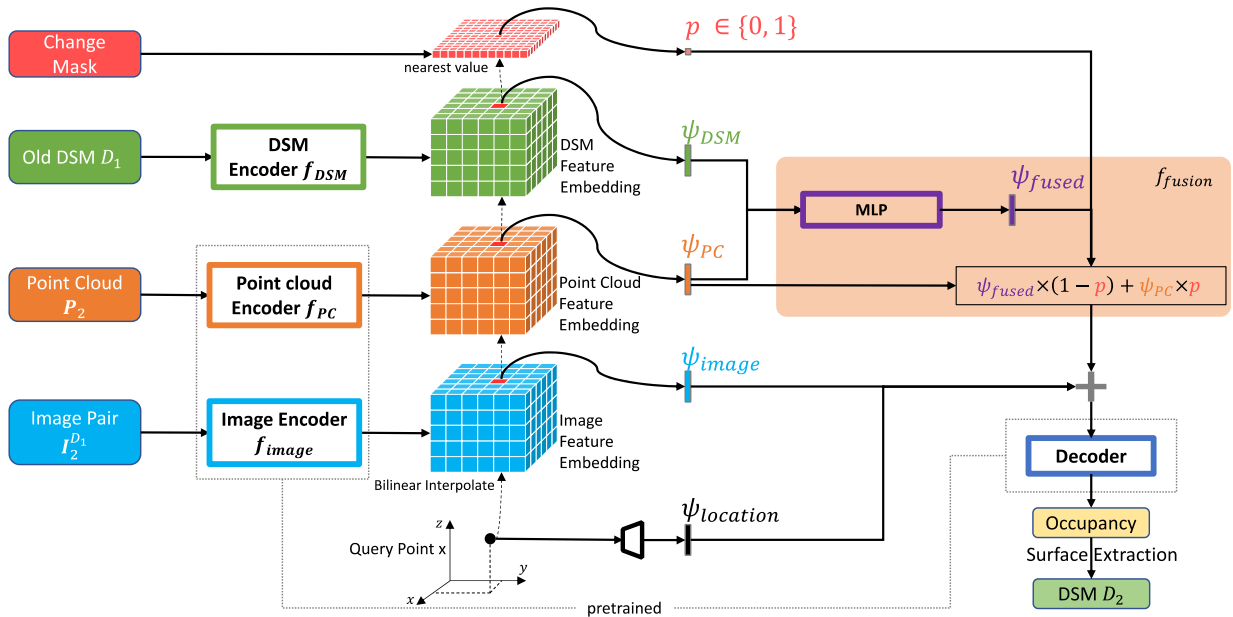


Fig. 2: Network architecture of the DSM updater



### 3 Experimental Setup

**Dataset.** We evaluate the proposed method on a satellite image dataset of WorldView-2 and WorldView-3 over Zurich, acquired between 2014 and 2018. The reference DSM is rasterized from the publicly available LoD2 city model of Zurich in 2015 and 2018. As shown in Figure 3, the study area consists of three rectangular sub-regions, covering various kinds of changed and unchanged buildings. The areas of these sub-regions are 1.89km<sup>2</sup> (ZUR\_A), 1.84km<sup>2</sup> (ZUR\_D), and 0.73km<sup>2</sup> (ZUR\_C). By choosing 2017-09-01 and 2018-12-01 as time thresholds, as shown in Figure 4, we build up a typical map updating scenario, where we have several images from the past that represent the old state; and few images, collected recently within a short time interval, to capture the new state. The dataset is separated into three different geographical areas for training, validation, and testing.

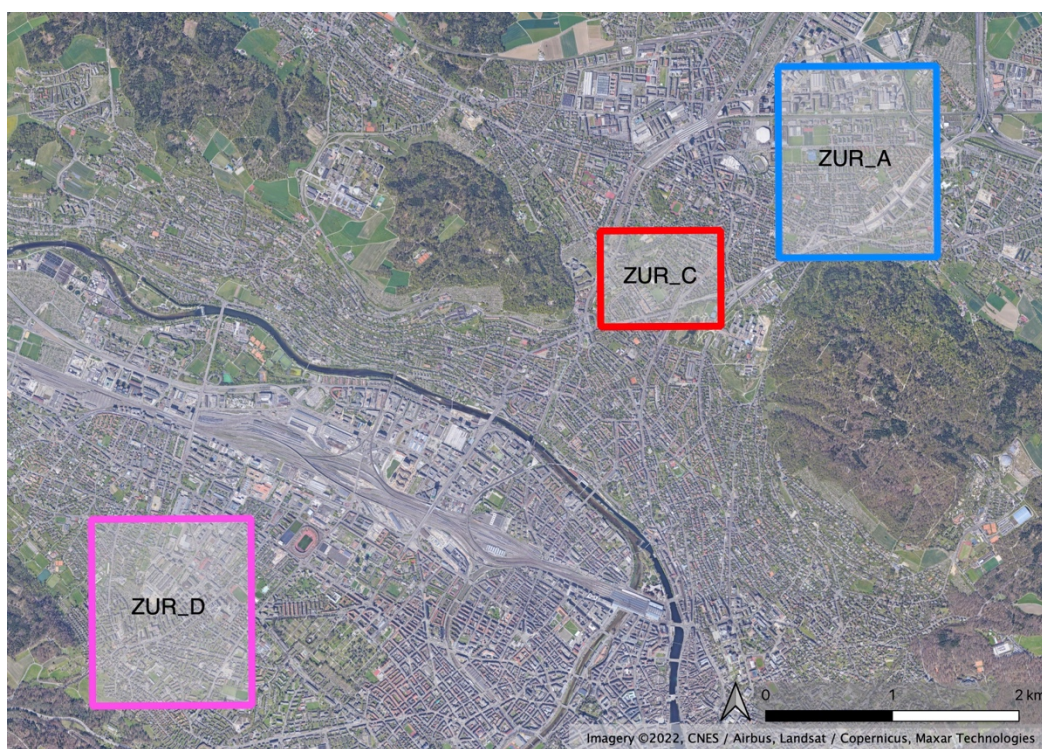


Fig. 3: Study area in Zurich. Three sub-regions with significant changes have been selected: ZUR\_A is training area, ZUR\_D is validation area, and ZUR\_C is test area.

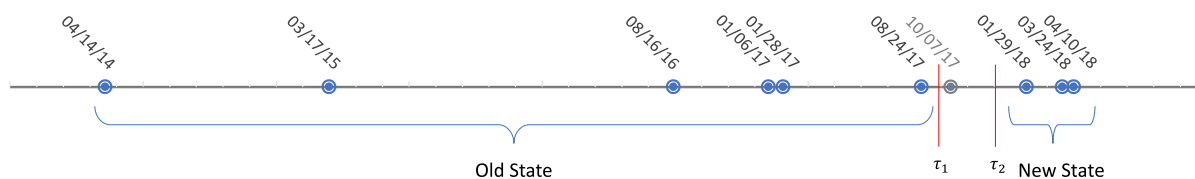


Fig. 4: Splitting of the image time series into data for the old and new states.

**Baseline.** We reconstruct the scene from scratch with only the new data by training the full *ImpliCity* network on the new data. This result represents the quality one can reach without taking into account any information about the old DSM state.

**Evaluation Metrics.** To quantify the quality of reconstructed and updated DSM, we calculate the mean absolute error (MAE) and the root mean square error (RMSE) over pixelwise deviations from the ground truth DSM.

**Implementation Details.** Our method is implemented in PyTorch and tested on a single NVIDIA GeForce RTX 2080Ti GPU. Training data is randomly sampled as patches with spatial dimension of  $64\text{m} \times 64\text{m}$  and randomly augmented by rotations of  $\alpha \in \{0^\circ, 90^\circ, 180^\circ, 270^\circ\}$  and flipping along the  $x$ - or  $y$ -axis. At inference time, we use a sliding window with 50% overlap to cover the whole validation or test area. Overlapping predictions are merged with linear blending.

## 4 Result and Discussion

**Quantitative Results.** Comparing the first two rows on Table 1 (i.e., old DSM vs. baseline), we see that both baseline reconstructions, using either only old data or only new data, have reasonably good quality. However, as expected the old data cannot correctly recover changed areas, whereas the new data alone suffers from low redundancy and yields worse overall performance. By comparing the 3<sup>rd</sup> and the 1<sup>st</sup> row, we see that leveraging the new data improves the MAE in changed area by  $\approx 50\%$ .

From the 3<sup>rd</sup> and the 2<sup>nd</sup> row we can see that the old DSM ensures the reconstruction quality in the unchanged area, and that the proposed selective DSM updating scheme outperforms the baselines in terms of overall statistics. The MAE drops by  $\approx 0.2\text{m}$  and the RMSE drops by  $\approx 0.3\text{m}$ .

Tab. 1: Quantitative comparison of reconstructed in the test area.

	Input state	Overall		Building		Terrain		Changed		Unchanged	
		MAE [m]	RMSE [m]	MAE [m]	RMSE [m]	MAE [m]	RMSE [m]	MAE [m]	RMSE [m]	MAE [m]	RMSE [m]
Input old DSM	Old	1.54	2.55	2.53	3.88	<b>1.23</b>	1.97	3.13	4.67	<b>1.50</b>	2.47
Baseline	New	1.71	2.69	2.45	3.87	1.48	2.22	<b>1.41</b>	<b>2.17</b>	1.71	2.71
Updated DSM	Old + New	<b>1.52</b>	<b>2.38</b>	<b>2.21</b>	<b>3.47</b>	1.31	<b>1.93</b>	1.59	2.63	1.52	<b>2.37</b>

**Qualitative Results.** Visual comparisons are shown in Figure 3. Firstly, we find that our method succeeds to update the content in changed areas from limited data, see white arrows in column (a) and (c). The demolished buildings are removed in the updated DSM, while the new buildings are reconstructed successfully.

Secondly, as pointed out by the violet arrows in column (a) and (b), our method correctly reproduces unchanged buildings that could not be reconstructed correctly only from the new data.

Moreover, as indicated by the red arrows in column (b) and (c), unchanged buildings that were missed in the old DSM can sometimes be recovered, too, when adding the new data. Moreover, the combined evidence tends to yield slightly sharper edges.

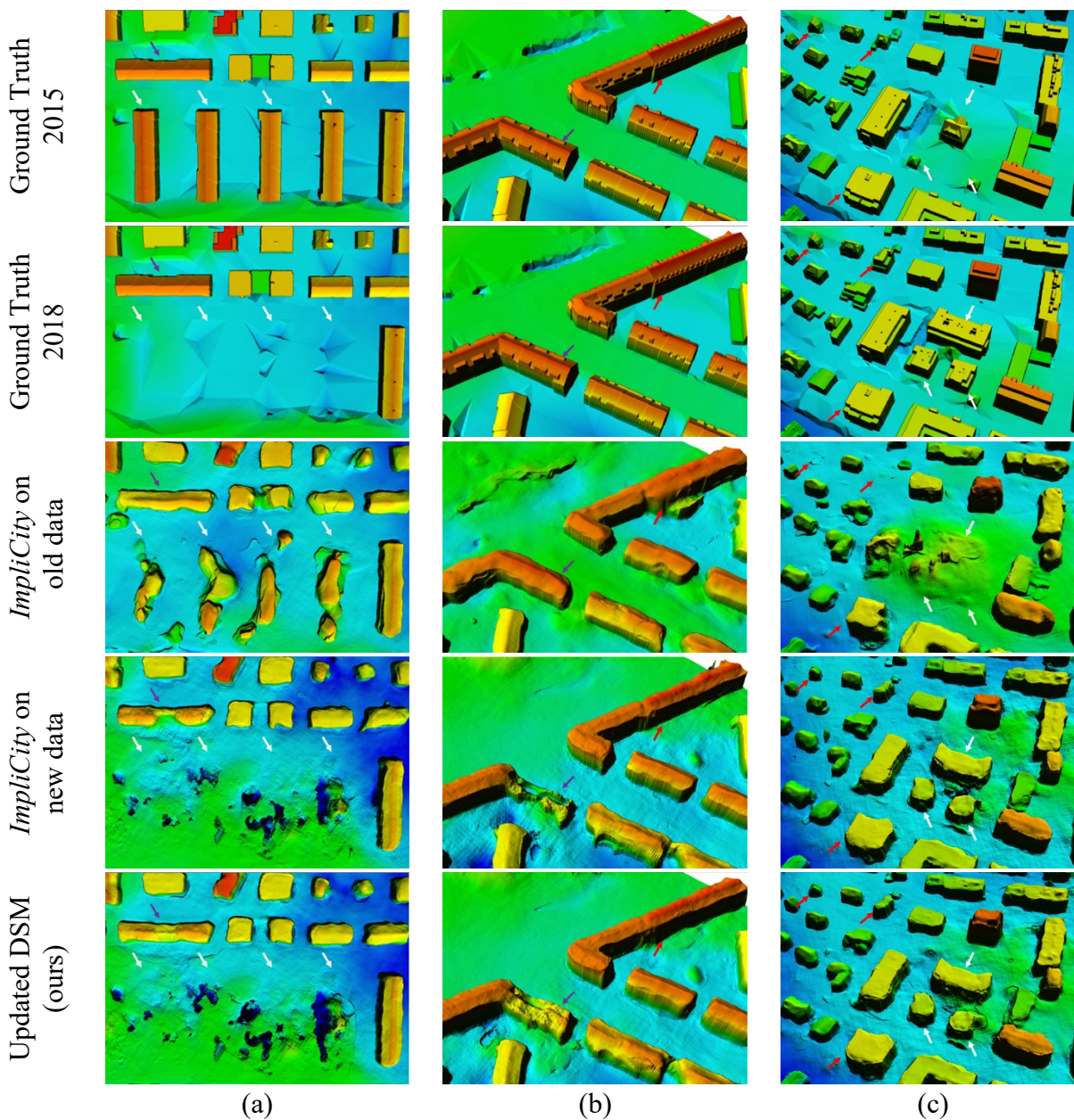


Fig. 3: Visual comparison of our result with the input DSM and the baseline in the test area. Heights are colored from blue to green to red. White arrows indicate the changed buildings; violet arrows indicate unchanged buildings that cannot be reconstructed as well from only new data; red arrows indicate unchanged buildings that are reconstructed better when using both old and new data.

## 5 Conclusion and Outlook

We have presented a DSM updating method for city scenes, using a DSM of an old, obsolete state together with satellite-based photogrammetric point clouds and ortho-photos generated with a small number of new images. The technical core of our method is a neural occupancy field. To the best of our knowledge, our work is the first learning-based approach for smart updating of city-scale DSMs based on satellite images. In summary, the conclusions of our study are:

1. *ImpliCity*, used from scratch with only little data that depict the new state, is able to reconstruct a DSM surprisingly well, still the quality is degraded in unchanged areas.
2. Leveraging the information captured in the old DSM together with the new data improves the reconstruction in the changed area without degrading unchanged areas. It is necessary to supply a reasonable change mask to guide the fusion, but that mask can be derived automatically from the available image data.
3. Suitably pretrained weights of the encoders and the decoder achieve comparable 3D quality in changed and unchanged areas, resulting in seamless fusion of old and new information.

In the wake of this exploratory study, potential improvements could be investigated in the future:

1. Reducing network complexity. The current method includes several fairly large modules. It would be desirable to reduce the parameter count while preserving performance.
2. Uncertainty guided data fusion. Introducing well-calibrated uncertainty estimation could provide additional information to better represent data dependencies.
3. Improving generalization. From an application point of view, it would be useful to ensure invariance w.r.t. varying imaging conditions, geographical areas, building styles, etc.

## 6 Acknowledgment.

This research is based upon work supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via Contract #2021-21040700001. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.

## 7 Bibliography

- BITTNER, K., D'ANGELO, P., KÖRNER, M., AND REINARTZ, P., 2018: DSM-to-LoD2: Spaceborne Stereo Digital Surface Model Refinement. *Remote Sensing*, **10**(12), 1926, <https://doi.org/10.3390/rs10121926>.
- DAUDT, R. C., LE SAUX, B. & BOULCH, A., 2018: Fully Convolutional Siamese Networks for Change Detection. *Proc. ICIP*, 4063-4067, <https://doi.org/10.1109/ICIP.2018.8451652>.
- DE FRANCHIS, C., MEINHARDT-LLOPIS, E., MICHEL, J., MOREL, J.-M. & FACCILOLO, G., 2014: An Automatic and Modular Stereo Pipeline for Pushbroom Images. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, **II-3**, 49-56, <https://doi.org/10.5194/isprsannals-II-3-49-2014>.
- PENG, S., NIEMEYER, M., MESCHEDER, L., POLLEFEYS, M. & GEIGER, A., 2020: Convolutional Occupancy Networks. *Computer Vision – ECCV 2020*, Vedaldi, A., Bischof, H., Brox, T. & Frahm, JM. (eds), *Lecture Notes in Computer Science*, **12348**, 523-540, Springer, Cham, [https://doi.org/10.1007/978-3-030-58580-8\\_31](https://doi.org/10.1007/978-3-030-58580-8_31).
- ROTHERMEL, M., WENZEL, K., FRITSCH, D. & HAALA, N., 2012: SURE: Photogrammetric Surface Reconstruction from Imagery. *Proc. LC3D Workshop*.

- STUCKER, C., KE, B., YUE, Y., HUANG, S., ARMENI, I. & SCHINDLER, K., 2022: ImpliCity: City Modeling from Satellite Images with Deep Implicit Occupancy Fields. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., V-2-2022, 193-201, <https://doi.org/10.5194/isprs-annals-V-2-2022-193-2022>.
- STUCKER, C. & SCHINDLER, K., 2022: ResDepth: A Deep Residual Prior for 3D Reconstruction from High-resolution Satellite Images. ISPRS Journal of Photogrammetry and Remote Sensing, **183**, 560-580, <https://doi.org/10.1016/j.isprsjprs.2021.11.009>.
- YOUSSEFI, D., MICHEL, J., SARRAZIN, E., BUFFE, F., COURNET, M., DELVIT, J.-M., L'HELGUEN, C., MELET, O., EMILIEN, A. & BOSMAN, J., 2020: CARS: A Photogrammetry Pipeline Using Dask Graphs to Construct a Global 3D Model. IGARSS 2020 - IEEE International Geoscience and Remote Sensing Symposium, 453-456, <https://doi.org/10.1109/IGARSS39084.2020.9324020>.

# Improving True Ortho Quality by Removing Moving Vehicles

HANNES NÜBEL<sup>1</sup> & PATRICK TUTZAUER<sup>2</sup>

*Abstract: Geometrically accurate representations like True Orthophotos (TOs) and textured meshes generated from aerial imagery are utilized progressively in many fields. Because of the growing interest in such products, these are optimized on many fronts like geometry, texturing, or processing time. An important customer need regarding the TO is the appropriate handling of moving vehicles. Apart from the ability to remove such objects completely, it is especially desired to prevent possible artifacts that are caused by a blending of varying textures from different images – an effect referred to as ghost cars. This paper describes an approach to remove moving vehicles from TOs by masking them in the aerial imagery with help from the information of depth images using a Convolutional Neural Network (CNN). Those masks are then deployed for the texturing of the TO. The approach proved to be effective with ~90% of the moving vehicles detected within the testing images and results in TOs with drastically removed moving vehicles and artifacts.*

## 1 Introduction

Capturing aerial imagery is a time-efficient method to map residential areas in a short time frame. However, because of the perspective projection of the image and the varying heights of the captured surface, they do not represent a map-like regular grid on the earth's surface. By means of a standard photogrammetric workflow, multiple overlapping images, initial positions, and orientations can be utilized to compute and refine 3D points and orientations, performing a bundle adjustment (KRAUS 2011). That information can then be used to map the texture information from the images onto the specified regular grid, creating an orthophoto. To be able to calculate the depth of corresponding points, they must be visible in at least two images.

One assumption in this process is that all captured objects are steady while taking the series of images. While geometric reconstruction from multiple views works reliably in most parts, texturing can pose difficulties. To get a consistent texture in the resulting product, it is common to use the color information out of multiple images for a single pixel in the orthophoto. If the information is taken from only one image, the results can get patchy, because of the different lighting conditions and viewing angles in the respective images. If the color of a pixel in the orthophoto is computed as the mean of the contributing pixels in multiple images, it is possible that the texture of a moving object - which is never at the very same position in multiple images - is blended with the surface. This can lead to effects like the so-called ghost cars shown in Fig. 1. This paper demonstrates the methodology and results of an approach that is able to remove such artifacts by detecting moving vehicles in the input imagery and masking them during the generation of the TO.

---

<sup>1</sup> Universität Stuttgart, Institut für Photogrammetrie, Geschwister-Scholl-Straße 24D, D-70174 Stuttgart, E-Mail: hannue@gmx.de

<sup>2</sup> nFrames | Esri R&D Center Stuttgart, Kornbergstraße 36, D-70176 Stuttgart, E-Mail: ptutzauer@esri.com



Fig. 1: Ghost cars induced by blending the texture of vehicles and street from different aerial images.

## 2 Related Work

One of the main scopes of the proposed approach is the detection of cars in images. This task has been discussed in many scientific papers and is now frequently implemented in the form of CNNs. For car detection, it is often required to detect each as a single object (e.g. to derive the number of cars). Therefore, instead of semantic segmentation, where each pixel in an image obtains a label, this work relies on instance segmentation to derive a mask and class label for each detected object. AUDEBERT et al. (2017) are using a three-step approach to first perform a semantic segmentation, then derive vehicle objects by searching for corresponding connected pixels, and finally, classify the instances into different types of vehicles. In the same year, Mask R-CNN (HE et al. 2017) was published, which was presenting an approach for a single CNN performing instance segmentation with masks on a pixel level. Mask R-CNN is also used as a base to solve the vehicle detection task in this study.

In contrast to other studies, this approach is utilizing aerial imagery together with photogrammetric data in the form of depth images, to only detect moving vehicles. The position and extent of each detected object are used during the texturing of the TO to exclude those image regions, which otherwise could lead to artifacts caused by the blending of inconsistent textures. Furthermore, it prevents texture in the TO that does not represent the reconstructed geometry in the Digital Surface Model (DSM).

### 3 Detection and Masking of Moving Vehicles

To be able to exclude image regions that show moving vehicles from the TO generation, it is necessary to create masks that indicate which pixels represent moving vehicles and should therefore be ignored. The mask for an aerial image is derived using an adapted and trained version of the Mask R-CNN (Fig. 2), which is performing instance segmentation to derive a semantic mask for each moving vehicle in the corresponding image. The CNN consists of a backbone network, which produces feature maps. To identify areas that could contain the desired objects, anchor boxes are learned using a Region Proposal Network. The pixels from the feature map of those anchor boxes are then passed on to determine the class, as well as a pixel-accurate mask of the respective object within the image section. Since aerial images are typically taken by large format cameras, and therefore are very large, they get tiled before being fed into the adapted Mask R-CNN. This also allows to keep a consistent input image size without resampling and stretching the original image.

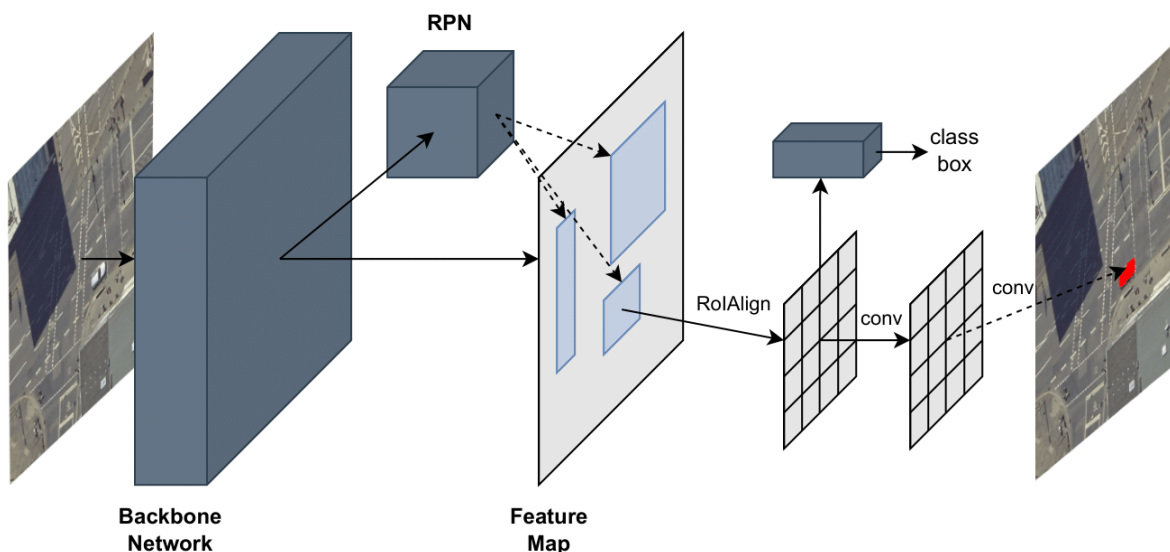


Fig. 2: Mask R-CNN framework (adapted from HE et al. 2017).

In addition to the color images captured by the camera, the respective depth images are used as a second input to the CNN. Although it is possible to check if certain objects are stationary in a series of images, this information is not contained in a single color image. The reconstruction pipeline SURE (ROTHERMEL et al. 2012) is producing depth images as an intermediate product from which such information about stationariness can be derived implicitly. A depth image depicts the distance of the ground points to the camera for the matched pixels in the base image. However, these depth images cannot be completely filled with values. Depth values can only be derived in areas where the used images are overlapping, so that corresponding pixels in stereo image pairs can be matched. Furthermore, for some image areas it is not possible to derive depth values, because the respective areas in the object space are occluded for the stereo image. Pixels of objects that are moving between the capturing of images also cannot be matched, because the pixels in this location show a different object than the pixels of another image mapping the same region. In



the case of moving cars, this means that an image that contains a car at a certain position will show the surface of the road for the same region in a different image. This results in holes in the depth image when a moving vehicle is captured. Moving vehicles can therefore be detected for each image when using the color information from an RGB image in conjunction with the depth image to tell if they are stationary (Fig. 3). The image stack consisting of RGB and depth image is fed to the CNN, which produces masks for the moving vehicles. Those are then used in the TO texturing process, to exclude the pixels in the color images and use color information from different views instead.

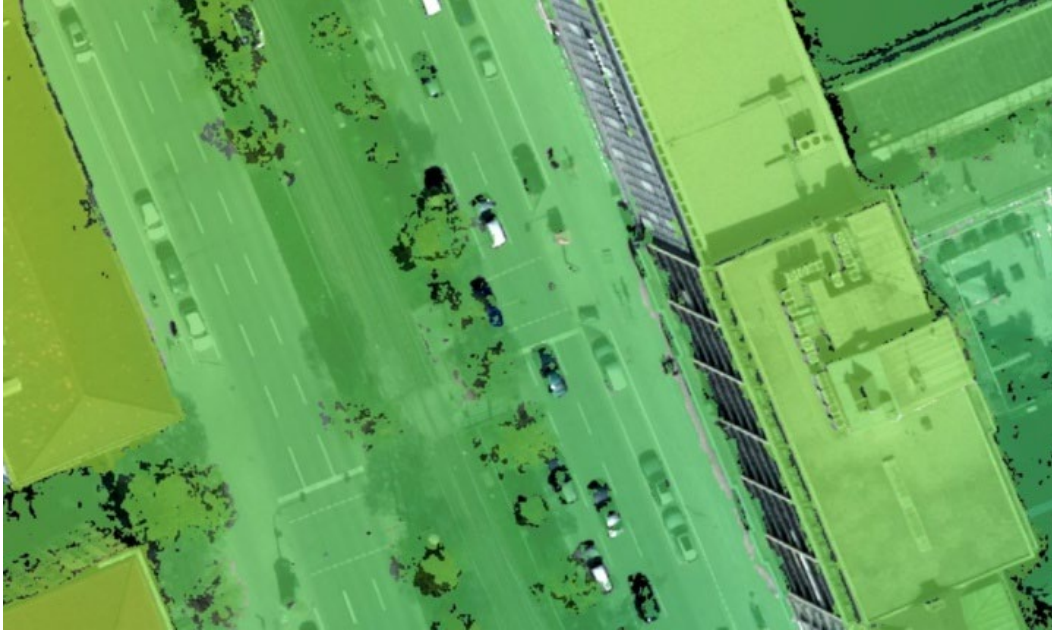


Fig. 3: RGB with transparent overlaid color-coded depth image (NaN values for depth image shown as no data) showing the difference between moving and stationary vehicles in the depth image.

To be able to train the CNN to identify moving vehicles, a large amount of training data is necessary. To create ground truth data for all aerial images, an existing car detection CNN was used to create masks for all cars, relying only on RGB information. To generate as many detections as possible, the probability threshold was set very low. The resulting detections were then automatically reduced by the cars that were clearly stationary because of a complete coverage of depth values. While strongly reducing manual labor, it was still necessary to delete false positives and add undiscovered moving vehicles by hand. Of the thereby created data 80% was used for training, while 20% was held out for final testing evaluations.

The process to create an improved TO without ghost cars was implemented as an extension to the reconstruction software SURE in a proof-of-concept fashion. It can be executed as one pipeline and does not need further manual interference, apart from setting up the standard inputs and parameters. After SURE has completed the dense matching and written out the depth maps for each base image, those are used together with the input imagery as input for the adapted Mask R-CNN, which is predicting pixel-accurate masks for all moving vehicles. When the DSM was processed, the moving vehicle masks for each image are used as a secondary input to the TO texturing process (Fig. 4).

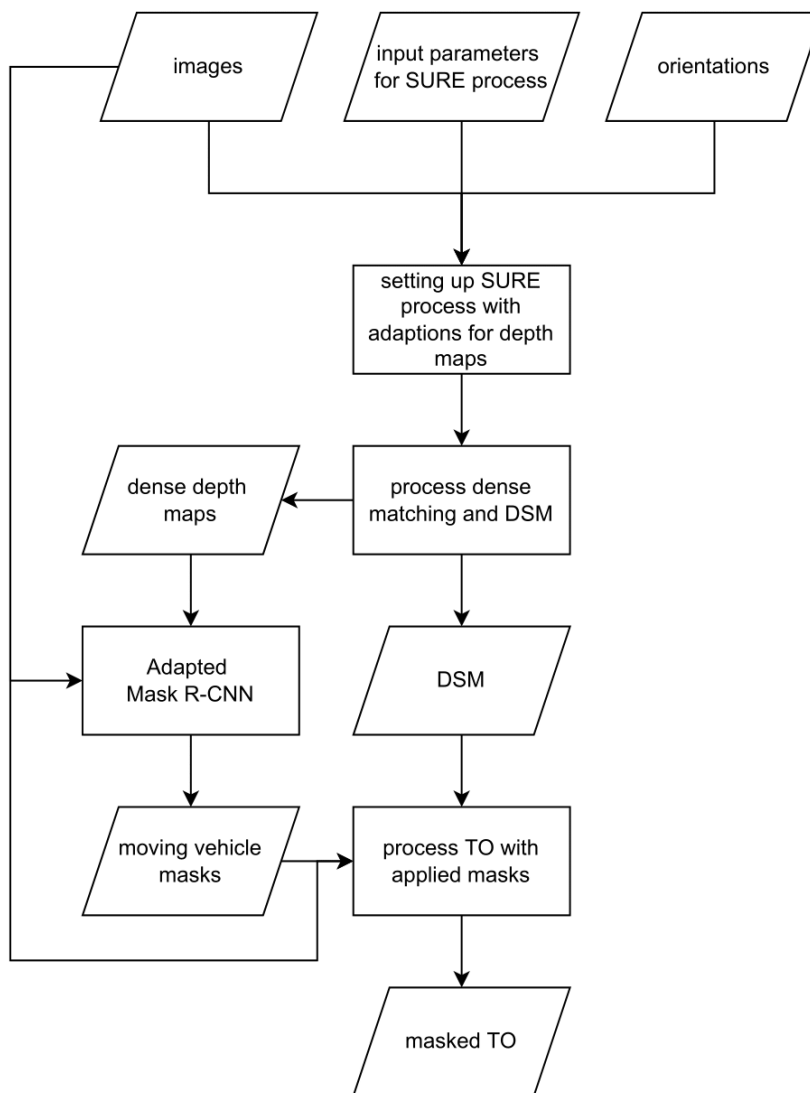
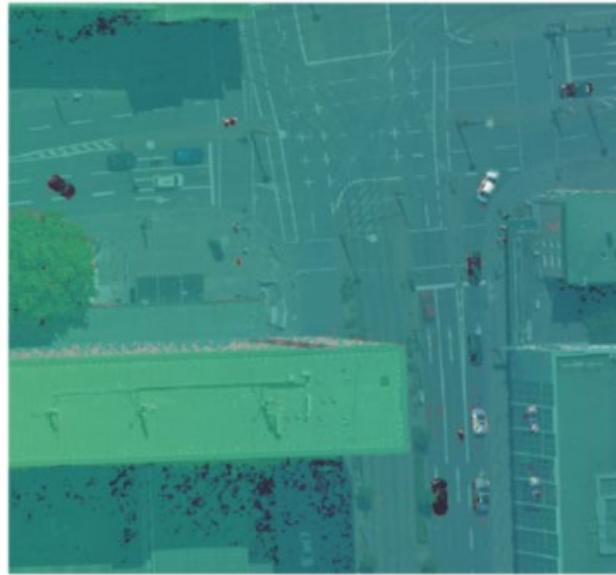


Fig. 4: Flowchart of pipeline integrating Mask R-CNN and SURE.

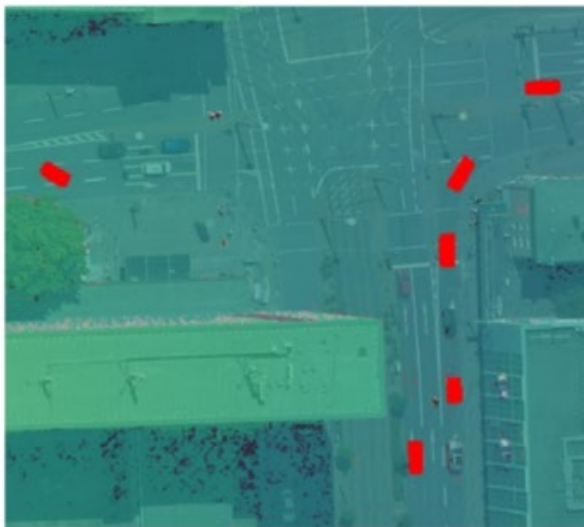
## 4 Results

### 4.1 Aerial Images with Moving Vehicle Masks

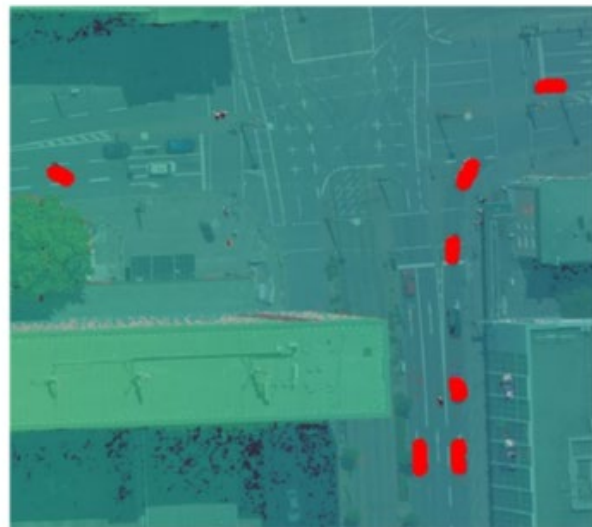
Fig. 5 shows the successful detection of cars and indicates that the CNN is also able to distinguish between moving and stationary. It also highlights a very challenging issue when vehicles are not constantly and/or only slowly moving. In this case, some points on the vehicle can be matched between images, resulting in an inconsistent set of depth values, as can be seen for the cars coming from the bottom and turning right. This is not only difficult when trying to predict an unseen image, but already when creating the ground truth for the training data. It has to be decided based on color and depth image, whether a vehicle is moving or not. If potentially moving objects are excluded from the ground truth masks, those could lead to unwanted effects in the TO if such vehicles do not get detected. On the other hand, if they are included in the training, many false positives might be detected for data sets or areas with sparsely filled depth images.



(a) RGB + depth image



(b) ground truth objects



(c) predictions

Fig. 5: Comparison of ground truth and prediction with moving, stationary, and partly moving vehicles.

If depth images have many invalid pixels, the precision of the CNN prediction drops because it is not possible to distinguish areas with moving objects from areas with insufficient matching caused by other issues like low overlap in the depth images. Since moving objects are one of the main contributors to missing depth information, completeness of depth maps is preferred over highest possible accuracy in the parameterization of this work.

## 4.2 Masked True Orthophotos

Fig. 6 shows segments of a TO with moving vehicles and the standard texturing approach on the left. The masked version, excluding the corresponding pixels in the aerial images from the texturing of the TO, is shown on the right. It can be observed that moving vehicles are effectively removed from the TO while the stationary ones, at traffic lights or parking spaces, retain their visualization. The approach is removing moving vehicles that are displayed completely in the reference, as well as ones that are blended, cut, textured multiple times and combinations of such effects that are especially unpleasant.



Fig. 6: TO comparison reference (left) and masked (right).

While detecting many moving vehicles in aerial images, the approach still relies on sufficient texture for those regions in other images. If the images have different lighting conditions, or never show the surface of the road (e.g. because of extreme traffic), data gaps or inconsistent texture for the masked areas in the TO can occur.

Figure 7 shows another data set with a very crowded road segment. The approach proves to be very effective in such challenging scenarios as well. All of the flowing traffic gets removed and on the left side replaced by matching textures from different images. On the right, there are some artifacts in the areas where cars got removed. This occurs because some of the images were captured with a significant temporal shift, causing the shadows of the larger buildings to move. Having no texture information for the areas of the moving vehicles, the texturing algorithm has to rely on texture information from other images, which in this case is inconsistent with the lighting conditions of their surroundings.



Fig. 7: Comparison of reference (top) and masked (bottom) TO with varying lighting conditions for different images on the right of the excerpt.

## 5 Conclusion

The approach described in this work proved to fulfill the need of removing moving vehicles from TOs for a high percentage of objects, creating a better view of the road networks, as well as mitigating many undesirable effects like ghost cars coming from an inconsistent blending of textures from moving vehicles. Since this approach is not changing the actual texturing procedure, the illustration of areas that do not have to be masked is staying the same. In addition, areas with masked moving vehicles mostly blend in with their surroundings since the texturing is performed equally by simply excluding the specified area from an image. However, the similarity is of course dependent on the views available for the corresponding area.

Two main challenges were encountered with the proposed approach. One is the necessity for dense depth maps to be able to reliably distinguish moving from stationary vehicles. The second is the handling of vehicles that cannot be labeled explicitly as moving or stationary for all images in which they are captured. This may lead to vehicles not being removed, i.e. appearing as for the reference TO, although they might cause artifacts. Nevertheless, the majority of moving vehicles are being detected and removed especially for flowing traffic, creating a more appealing end result.

To further improve the precision of the CNN it would be possible to include other byproducts from the reconstruction pipeline that provide additional information about the matching of points. Apart from that, the detections could be classified into different kinds of moving vehicles, to be able to reliably detect vehicles of different shapes and sizes.

## 6 References

- AUDEBERT, N., LE SAUX, B. & LEFÈVRE, S., 2017: Segment-before-detect: Vehicle detection and classification through semantic segmentation of aerial images. *Remote Sensing*, **9**(4), 368.
- GOODFELLOW, I., BENGIO, Y. & COURVILLE, A., 2016: *Deep Learning*. MIT press.
- HE, K., GKIOXARI, G., DOLLÁR, P. & GIRSHICK, R., 2017: Mask r-cnn. *Proceedings of the IEEE international conference on computer vision*, 2961-2969.
- KRAUS, K., 2007: *Photogrammetry: Geometry from Images and Laser Scans*. Berlin, Boston: De Gruyter, <https://doi.org/10.1515/9783110892871>.
- REN, S., HE, K., GIRSHICK, R. & SUN, J., 2015: Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, **28**.
- ROTHERMEL, M., WENZEL, K., FRITSCH, D. & HAALA, N., 2012: SURE: Photogrammetric surface reconstruction from imagery. *Proceedings LC3D Workshop*, Berlin.

# Real-Time Tracking and 3D Dense Reconstruction Based on ORB-SLAM3 Extensions using a Depth Camera

JIWEI HOU<sup>1</sup>, MONA GOEBEL<sup>1</sup> & DOROTA IWASZCZUK<sup>1</sup>

*Abstract: Simultaneous localization and mapping (SLAM) is widely used for robot perception of the environment. It supports the robot in determining its own position as well as the position of surrounding objects. Due to the low cost and the intuitive approach to capture the environment, there are many visual SLAM (vSLAM) systems that have been published in the past decades. In this work, we will evaluate the SLAM algorithm ORB-SLAM3. ORB-SLAM3 is one of the best performing open source vSLAM algorithms, which can not only automatically estimate the exact position and pose of the camera, but also construct a 3D sparse point cloud map of the working area. However, the sparse 3D maps produced cannot meet the requirements of robots performing tasks such as obstacle avoidance, planning, and navigation autonomously. To improve the usability of the sparse maps extracted from ORB-SLAM3, this work investigates ways to increase the density of point clouds. For this, we reconstructed dense point cloud maps of 3D scenes using an extended ORB-SLAM3 mapping algorithm based on RGB-D images and camera poses. We tested our dense mapping system with the benchmark TUM RGB-D dataset published by the Technical University of Munich. Thereafter, we collected data with RealSense depth camera D455 and got a good real-time dense mapping result.*

## 1 Introduction

The problem of simultaneous localization and mapping (SLAM) in robotics has attracted many researchers working on problems in this field over the last few decades. Researchers have proposed a large number of SLAM systems, incorporating sensors, optimization algorithms, and map descriptions. All SLAM systems are designed to maintain system robustness, improve tracking accuracy, and achieve real-time performance. The availability of 3D maps is an important requirement for robots in different workspace conditions to autonomously perform multiple tasks including positioning, planning and navigation. Especially in complex and dynamic environments, it is critical for robots to quickly generate and maintain 3D maps through on-board sensors. For example, self-driving vehicles require high precision real-time maps to avoid obstacles and navigate safely in the complex real world.

Today's common SLAM systems include vSLAM and LiDAR SLAM. vSLAM uses cameras as its primary sensor, tracking the pose of the sensor while creating a map of the environment (FUENTES-PACHECO et al. 2015). LiDAR-based SLAM system uses laser sensors to generate a 3D map of its environment. Due to the low cost and the intuitive approach to create a 3D map, there have been many vSLAM systems published in the past decades. Representative examples include PTAM (KLEIN et al. 2007), LSD-SLAM (ENGEL et al. 2013), SVO (FORSTER et al. 2014), RGB-D

---

<sup>1</sup> Technical University of Darmstadt, Department of Civil and Environmental Engineering Sciences, Remote Sensing and Image analysis, Franziska-Braun-Str. 7, D-64287 Darmstadt.  
E-Mail: [jiwei.hou, mona.goebel, dorota.iwaszczuk]@tu-darmstadt.de

SLAM (ENDRES et al. 2014), ORB-SLAM (MUR-ARTAL et al. 2015) and ORB-SLAM3 (CAMPOS et al. 2021).

Among the many vSLAM solutions, ORB-SLAM is one of the traditional feature point-based SLAM algorithm. With the release of ORB-SLAM, ORB-SLAM2 (MUR-ARTAL et al. 2017) and ORB-SLAM3, ORB-based SLAM systems have been continuously updated and improved in the past few years. Especially ORB-SLAM3 has become one of the best performing feature-based SLAM system that operates in real time, both indoors and outdoors. ORB-SLAM algorithms are lightweight and can therefore be run on CPU hosts. However, ORB-SLAM aims at long-term and globally consistent localization rather than building the most detailed dense reconstruction. The previous work of ORB-SLAM2 reconstructed dense point clouds, but ORB-SLAM3 did not. Furthermore, the dense mapping code in ORB-SLAM2 was not published open-source. We apply ORB-SLAM3 in 3D indoor mapping and modelling, where a depth camera is used to scan the room. Therefore, it is very important to obtain a 3D dense point clouds reconstruction of indoor space. In this paper, we will focus on the mapping part of SLAM and investigate ways to increase the density of point clouds. Based on ORB-SLAM3, we extend the sparse map constructed by the original system to a dense point cloud map using RGB-D images and camera poses. Thereafter, we collected data with the RealSense depth camera D455. Our results are then compared with the TUM RGB-D dataset published by the Technical University of Munich (STURM et al. 2012). Finally, we used the OctoMap library (HORNUNG et al. 2013) to reconstruct an efficient probabilistic octree map for robotic applications.

The main contents of this paper are as follows: We discuss background and related work in Section 2, describe our system and method in Section 3. Lastly, in Section 4, we present the RGB-D dense mapping results and evaluation, and draw a conclusion in Section 5.

## 2 Background and Related work

We will first highlight the progress in the field of vSLAM research. Thereafter, ORBSLAM3 is explained in more detail. We close this section with an overview of publications using Red-Green-Blue-Depth (RGB-D) images for vSLAM and introduce the well-known Point Cloud Library.

### 2.1 Visual SLAM

As we mentioned in the introduction, researchers have developed many SLAM solutions, most of which are open source on GitHub. We selected some typical ones as reference, especially focus on 3D dense reconstruction based on RGB-D sensors. GEORG KLEIN and DAVID MURRAY proposed the PTAM algorithm (2007), which splits tracking and mapping into two parallel threads. The system can easily track across multiple scales and provide tracking quality suitable for small workspace augmented reality (AR) applications. DTAM (GANAI et al. 2012) is a system for real-time camera tracking and reconstruction, it does not rely on feature extraction, but on a dense per-pixel approach. The DTAM algorithms are highly parallelizable throughout and rely on GPU to achieve real-time performance. LSD-SLAM (ENGEL et al. 2013) is a feature-less monocular SLAM algorithm which runs in real-time on a CPU. The algorithm changes the pixel selection to make it suitable for larger scale scenarios. However, as it is a feature-less method based on the



assumption of grayscale invariance, its robustness and accuracy may be affected by unmodeled behaviors such as lens vignetting and drastic changes in illumination. RGB-D SLAM (ENDRES et al. 2014) system uses Random Sample Consensus (RANSAC) to estimate the transformations between associated key points and optimizes the pose graph using non-linear optimization. This takes advantage of the dense color and depth images provided by RGB-D cameras to estimate camera pose and 3D environment construction. RGB-D SLAM can robustly handle challenging scenarios, such as fast camera movements and feature-poor environments, while being fast enough to operate online. ZHANG et al. (2022) proposed a comprehensive visual SLAM system that extends the application of ORB-SLAM3, which realized 3D dense reconstruction with both RGB-D and stereo cameras. Although they have not made their system open source, we did get a lot of inspiration from their paper.

## 2.2 ORB-SLAM3

ORB-SLAM3 (CAMPOS et al. 2021) is a visual SLAM algorithm that supports multiple cameras. This algorithm optimizes several aspects such as map initialization, relocation, closed-loop detection, key frame selection, map construction, sensor support, and ultimately provides excellent performance in terms of operation speed, tracking effect and composition accuracy.

The algorithm of ORB-SLAM3 includes three main threads: tracking thread, local mapping thread, loop detection and map fusion. The tracking thread performs rough processing of the input data, including feature point extraction, frame matching, key frame filtering, and converts frames and map points into nodes and edges to provide reliable initial values for subsequent threads. The local mapping thread further filters and optimizes the key frame data, and uses G2O (KÜMMERLE et al. 2011) to optimize the pose relationship between frames and map points. Loop detection and map fusion improve the drift error and multi-map management.

## 2.3 RGB-D cameras

RGB-D cameras are active image sensors that can not only collect color information, but also calculate depth using these stereo images. Some have additional LiDAR sensors included, to measure distance in space. In this paper, we use Intel® RealSense™ Depth Camera D455, as shown in Fig.1a. This camera can acquire RGB three channels color images and the corresponding depth data, as in Fig.1b and 1c. It uses active Infrared (IR) stereo vision technology to measure depth information with a left and right image sensor and an optional IR projector. The detection distance is between 0.52 m and 6 m. IR projector projects invisible static infrared patterns to improve depth accuracy in low texture scenes. The left and right image sensors capture the scene and send image data to a depth imaging (vision) processor that calculates the depth value of each pixel in the image by associating points between the left image and the right image. Moreover, the RGB sensor includes a global shutter and is matched to the depth Field of View (FOV), improving not only the quality of RGB images but also the correspondence between depth and RGB images. Another benefit is that the D455 camera supports self-calibration without the need for specialized targets. Intel® RealSense™ on-chip calibration allows accurate inspection of the system to ensure the sensor is operating in the optimum range. This makes the D455 even more convenient to calibrate and use.

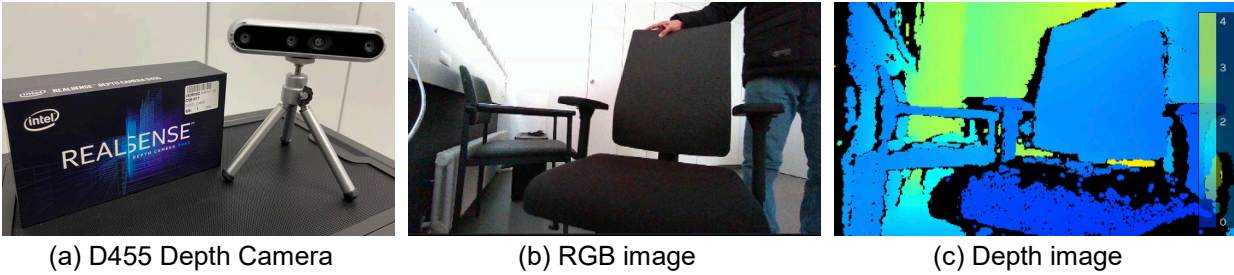


Fig. 1: The Intel® RealSense™ Depth Camera D455 (a), with an example RGB image (b) and its corresponding depth image (c)

## 2.4 Point Cloud Library (PCL)

PCL (RUSU et al. 2011) is a powerful cross-platform open source C++ programming library developed based on previous research on point clouds. It started as an open-source project under the Robot Operating System (ROS), which is developed and maintained by Dr. RADU and others from the Technical University of Munich (TUM) as well as researchers from the Stanford University. ROS is mainly for robotics research applications. Furthermore, PCL implements a large number of general algorithms and efficient data structures related to point clouds such as acquisition, filtering, segmentation, alignment, retrieval, feature extraction, identification, tracking, surface reconstruction and visualization. It supports multiple operating system platforms and can run on Windows, Linux, Android, Mac OS X, and some embedded real-time systems.

## 3 Methods

In this paper, we reconstruct point cloud maps based on ORB-SLAM3 system. The pipeline of this work is shown in Fig. 2. ORB-SLAM3 algorithm performs high precision estimation of camera pose and sparse maps simultaneously. In the program, we create a single thread to extend this work, through invoking every single keyframe from ORB-SLAM3 directly, and combining it with the acquired optimized camera poses and RGB-D information. Using RGB images and depth images, we can acquire point clouds per session. Finally, we merge point clouds based on camera poses and compose the map. To improve the quality and accuracy of the point cloud reconstruction, we use a two-layer filtering method: (1) statistical filtering in the local map to remove the outlier points, and (2) voxel filtering in the global map to downsample which reduces the number of points in the point cloud. Thereby, the shape characteristics are preserved, and less memory is used without serious distortions. Furthermore, an octree map was constructed which can be applied to robot obstacle avoidance, navigation and interactive manipulation.

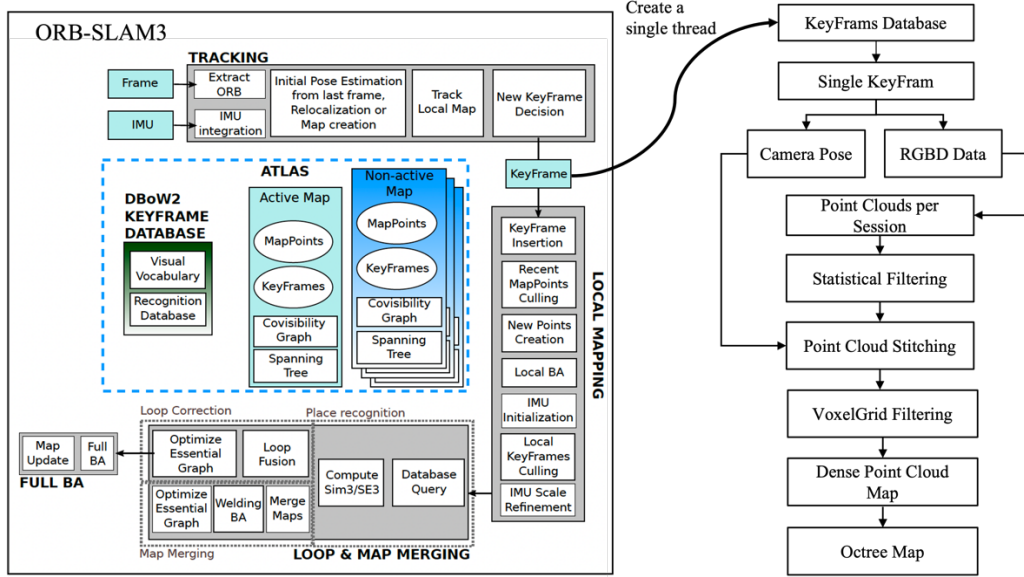


Fig. 2: Pipeline of dense point cloud map reconstruction based on ORB-SLAM3 (Campos et al. 2021)

### 3.1 Statistical filtering

One of the more novel methods of point cloud denoising in PCL is the StatisticalOutlierRemoval filter, which calculates the distribution of distances from each point to its neighbor in the input data and obtains the average distance from each point to all its neighbor. The result is assumed to be a Gaussian distribution whose shape is determined by the mean  $\mu$  and standard deviation  $\sigma$ . Assume the coordinate of point  $P_n$  ( $X_n, Y_n, Z_n$ ), the distance from this point to any point  $P_m$  ( $X_m, Y_m, Z_m$ ) can be expressed as equation (3-1),

$$S_i = \sqrt{(X_n - X_m)^2 + (Y_n - Y_m)^2 + (Z_n - Z_m)^2} \quad . \quad (3-1)$$

The mean distance between each point to any points is calculated as in (3-2)

$$\mu = \frac{1}{n} \sum_{i=0}^n S_i \quad . \quad (3-2)$$

The standard deviation is given by formula (3-3)

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (S_i - \mu)^2} \quad . \quad (3-3)$$

Set the standard deviation multiplier as  $std$ , retain the point  $P_n$  when the mean distance between point  $P_n$  and adjacent  $k$  points is within the standard range ( $\mu - \sigma \times std$ ,  $\mu + \sigma \times std$ ), and if it is not within that range, it is defined as an outlier and removed from the point clouds dataset.

### 3.2 VoxelGrid filtering

The voxel filter allows for downsampling without destroying the geometry of the point cloud itself, but it does shift the position of the points. In addition, the voxel filter can remove a certain amount of noise and outliers. The main function is to downsample.

The principle of VoxelGrid filtering is to first calculate a cube that can easily enclose the input point cloud, and then split the large cube into smaller cubes according to a set resolution. For each of the points within the small cube, their centroid is calculated, and the coordinates of the centroid are used to approximate a number of points within the cube, thus reducing the amount of data. It is therefore often used for downsampling large amounts of data, especially as pre-processing step before, for example, the alignment or surface reconstruction. This can be a good way to increase the efficiency of the program. The calculation steps of VoxelGrid filtering are as follows.

The first step is to determine the edge length  $L$  of each unit voxel, the formula is (3-4)

$$L = \alpha \sqrt[3]{\frac{s}{n}} \quad (3-4)$$

$$n = \frac{N}{V} \quad (3-5)$$

where  $s$  is the scale factor,  $\alpha$  is the scale factor used to adjust the edge lengths of unit voxels, and  $n$  is the number of points per voxel, which can be calculated by (3-5).  $N$  represents the total number of points,  $V$  represents the volume of each unit voxel. By adjusting  $\alpha$ , the algorithm can dynamically adapt to the sparsity of each part of the point cloud. Using (3-4) and (3-5) we can get the relationship between the unit voxel edge length and the total number of points as

$$L = \alpha \sqrt[3]{\frac{s \times V}{N}} \quad (3-6)$$

The volume of the unit voxel is ,

$$V = L_x \times L_y \times L_z \quad (3-7)$$

where  $L_x$ ,  $L_y$ , and  $L_z$  represent the projections of unit voxels on the x, y, and z axes, respectively. We can then replace  $V$  in equation (3-6), resulting in

$$L = \alpha \sqrt[3]{\frac{s \times L_x \times L_y \times L_z}{N}} \quad (3-8)$$

The second step is to calculate the centroid coordinates per voxel, thus, finding the point cloud data which can represent the unit voxel.

$$\begin{cases} X_c = \sum_{i=1}^m \frac{x_i}{m} \\ Y_c = \sum_{i=1}^m \frac{y_i}{m} \\ Z_c = \sum_{i=1}^m \frac{z_i}{m} \end{cases} \quad (3-9)$$

In (3-9), where  $m$  is the number of points within the voxel.  $(x_i, y_i, z_i)$  are the coordinates of each point. The point cloud which is closest to the centroid coordinates is selected and retained in place of all points within the voxel to achieve VoxelGrid filtering downsampling.

## 4 Results and Discussion

To validate the performance of our proposed dense reconstruction system, we tested it with the benchmark RGB-D camera dataset from TUM as well as with the self-collected data from our office in real time scanning with a handheld D455 camera. All experiments were carried out on a computer with Ubuntu18.04 as operating system, AMD Epyc 7402p, 24-core processor, 256G RAM and NVIDIA RTX A4000 GPU.

### 4.1 Dense mapping results and evaluation

In this paper, we have chosen *fr2\_desk* and *fr3\_long\_office\_houhold* as input data. Fig. 4a and Fig. 4b are the experimental results displayed in the software CloudCompare. The well-defined and straight contours of the desks and chairs show the high accuracy alignment of our RGB-D dense mapping system. After estimating the camera trajectory of *fr2\_desk* and *fr3\_long\_office\_houhold*, the estimated result is saved in a local file. We then evaluate the error of the estimated trajectory by comparing it with ground truth data provided by the official website of TUM RGB-D dataset. There are different error metrics. The absolute trajectory error (ATE) is an ideal error metric for measuring the performance of vSLAM systems.

To evaluate the scanning performance of the system, we ran each data sequence three times and extracted the file size of the resulting point cloud and the number of points it contained. Tab. 1 shows the statistics of the point cloud data and the corresponding camera motion trajectory ATE. We can see that for the same TUM RGB-D sequence, the point number and file size in each point cloud data vary with each run. One reason for this could be the different number of keyframes chosen by the algorithm during point cloud reconstruction. Another reason could be the two-layer filtering process, effecting the final result.

After this first analysis, we used the ATE evaluation script provided by the official website of TUM RGB-D dataset, to compute the ATE of the estimated camera trajectories for each sequence with respect to the ground truth. Fig. 5 shows the ATE visualized for the selected sequences, and Tab. 2 shows the ATE root-mean-square error (RMSE), ATE mean, ATE median, ATE standard deviation (Std), ATE minimum and maximum in meters for the *fr2\_desk* and *fr3\_long\_office\_houhold* sequences. From these

Tab. 1: Statistics of point cloud maps for the selection of TUM RGB-D sequences. MB stands for megabytes

Datasets	<i>fr2_desk</i>			<i>fr3_long_office_houhold</i>		
	Number of points	1,320,810	1,322,930	1,369,868	1,618,853	1,592,130
Mean number of points	1,337,869			1,604,373		
Number of keyframes	229	236	238	284	274	273
Mean number of keyframes	234			277		
File size (MB)	21.1	21.2	21.9	25.9	25.5	25.6
Mean file size (MB)	21.40			26.67		

values, we show that the state of our RGB-D dense mapping system extended by ORB-SLAM3 is relatively stable without major drift.

Finally, this RGB-D point cloud dense reconstruction system is also successfully used for the real time mapping of the data collected with the handheld RealSense D455 camera. The computer reads the depth and color images from the D455 via USB 3.2 interface and publishes the image data as a ROS topic. We scanned the desk in the office and the room scene around the desk separately. The results are shown in Fig. 6a and Fig. 6b.

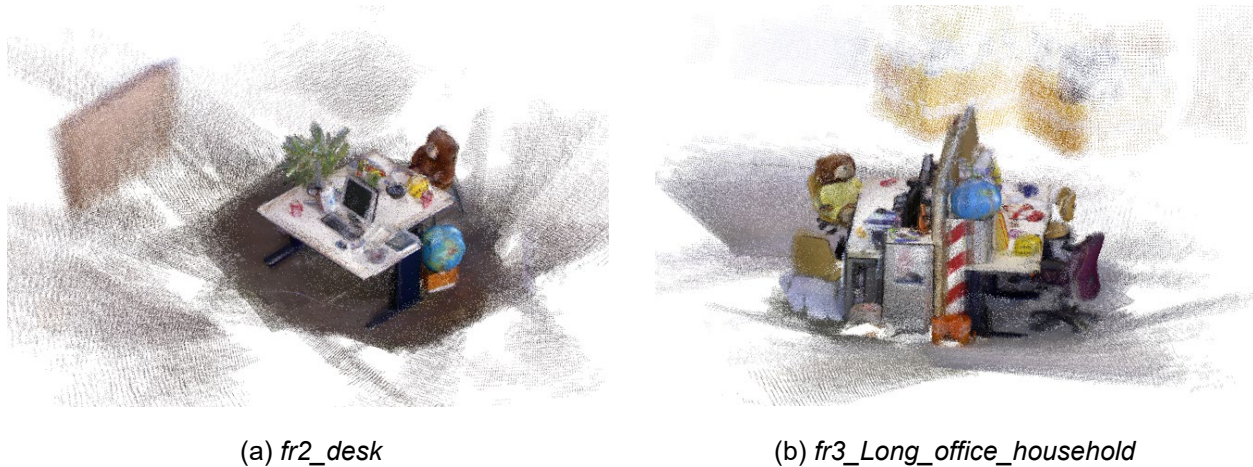


Fig. 4: The dense mapping result for the TUM RGB-D dataset

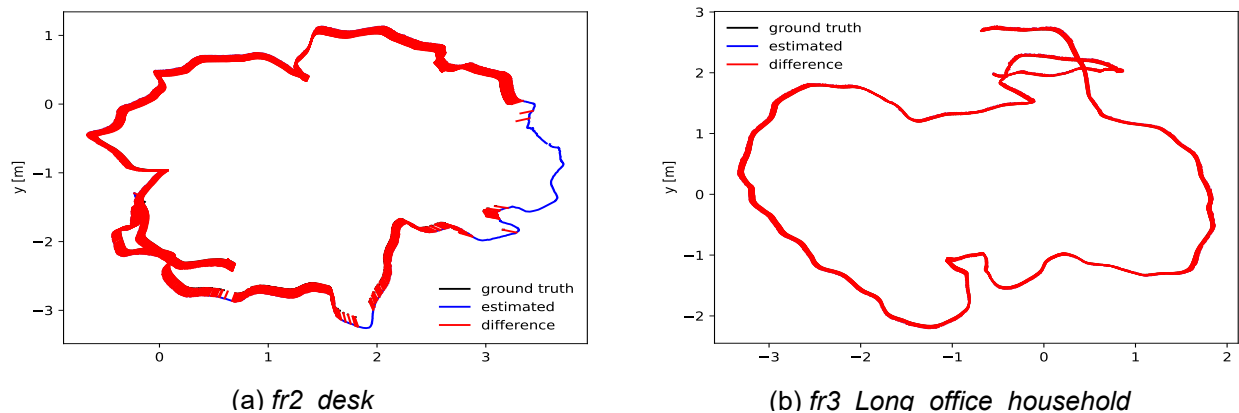


Fig. 5: The ATE results for fr2\_desk and fr3\_Long\_office\_household after comparing the estimated camera trajectory with ground truth

Tab. 2: Statistics of absolute translational error in meters for the selection of TUM RGB-D sequences.

Data	<i>fr2_desk</i>	<i>fr3_long_office_houhold</i>
RMSE	0.092832	0.031973
Mean	0.091270	0.030279
Median	0.093389	0.030554
Std	0.016960	0.010267
Min	0.049749	0.006144
Max	0.138055	0.055867

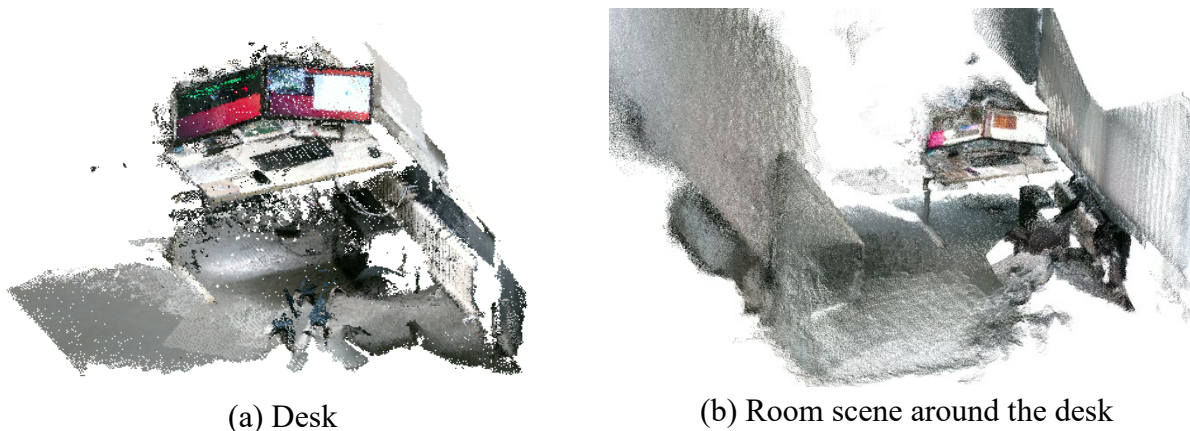


Fig. 6: The RGB-D dense mapping result for our dataset

#### 4.2 The Preliminary results of octrees

We used OctoMap library to convert the point cloud data into octrees data, the point cloud data is shown in Fig. 6a, the octrees results are shown in Fig. 7a and Fig. 7b, Fig. 7b is octrees map result with RGB information. The file size of the colored point cloud data was reduced from 7.3 megabytes to 5 kilobytes and 77 kilobytes (1 megabyte = 1024 kilobytes), the data volume was reduced by approx. 99.93% and 98.96%.

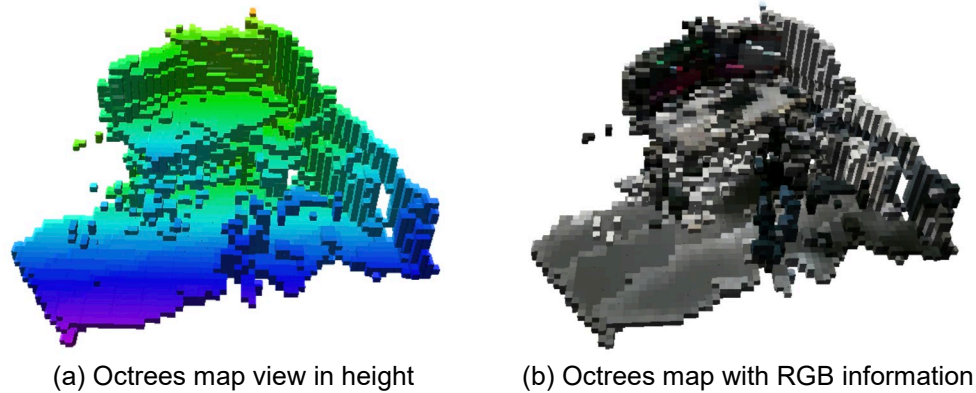


Fig. 7: Octomap reconstructed from desk in our office

## 5 Conclusion and future work

In this paper, we built a real-time 3D dense point cloud reconstruction system based on ORB-SLAM3, using the statistical and voxel filtering algorithms provided by PCL. Further, we successfully applied it on the self-collected Intel RealSense D455 camera data. This system was also tested on the TUM RGB-D. Shown through experiments indoors, the system can successfully generate indoor dense point clouds. Finally, an initial result of octrees mapping is displayed, which could be used in our future research for robotic automatic navigation.

As this is only a preliminary experimental result, we will further improve the robustness and performance of this system, such as adding filters and adding map loop detection to enhance the results. In addition, the octree map could be built in real time based on the point clouds, rather than the current offline generation. Moreover, we have only carried out real-time construction of the 3D point cloud in a small indoor room and have not yet tested it in a larger space, which is one of the next steps.

## 6 ACKNOWLEDGMENT

This research is supported by China Scholarship Council (CSC), Grant/Award Number: 202108130064.

## 7 References

- CAMPOS, C., ELVIRA, R., RODRIGUEZ, J. J. G., MONTIEL, J. M. M. & TARDOS, J. D., 2021: ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial, and Multimap SLAM. *IEEE Transactions on Robotics*, **37**(6), 1874-1890, <https://doi.org/10.1109/TRO.2021.3075644>.
- ENDRES, F., HESS, J., STURM, J., CREMERS, D. & BURGARD, W., 2014: 3-D Mapping with an RGB-D camera. *IEEE Transactions on Robotics*, **30**(1), 177-187, <https://doi.org/10.1109/TRO.2013.2279412>.
- ENGEL, J., STURM, J. & CREMERS, D., 2013: LSD-SLAM: Large-Scale Direct Monocular SLAM. *Proceedings of the IEEE International Conference on Computer Vision*, 1449-1456.



- FORSTER, C., PIZZOLI, M. & SCARAMUZZA, D., 2014: SVO : Fast Semi-Direct Monocular Visual Odometry. 2014 IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, China, 15-22. <https://doi.org/10.1109/ICRA.2014.6906584>.
- FUENTES-PACHECO, J., RUIZ-ASCENCIO, J. & RENDÓN-MANCHA, J. M., 2015: Visual simultaneous localization and mapping: a survey. *Artificial Intelligence Review*, **43**(1), 55-81, <https://doi.org/10.1007/s10462-012-9365-8>.
- GANAI, M., LEE, D. & GUPTA, A., 2012: DTAM: Dense Tracking and Mapping in Real-Time. *Richard*. 1, <https://doi.org/10.1145/2393596.2393650>.
- HÖRNING, A., WURM, K. M., BENNEWITZ, M., STACHNISS, C. & BURGARD, W., 2013: OctoMap: An efficient probabilistic 3D mapping framework based on octrees. *Autonomous Robots*, **34**(3), 189-206. <https://doi.org/10.1007/s10514-012-9321-0>.
- KLEIN, G. & MURRAY, D., 2007: Parallel tracking and mapping for small AR workspaces. 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, ISMAR, 225-234, <https://doi.org/10.1109/ISMAR.2007.4538852>.
- KÜMMERLE, R., GRISETTI, G., STRASDAT, H., KONOLIGE, K. & BURGARD, W., 2011: G2o: A general framework for graph optimization. *Proceedings - IEEE International Conference on Robotics and Automation*, 3607-3613, <https://doi.org/10.1109/ICRA.2011.5979949>.
- MUR-ARTAL, R., MONTIEL, J. M. M. & TARDOS, J. D., 2015: ORB-SLAM: A Versatile and Accurate Monocular SLAM System. *IEEE Transactions on Robotics*, **31**(5), 1147-1163, <https://doi.org/10.1109/TRO.2015.2463671>.
- MUR-ARTAL, R. & TARDOS, J. D., 2017: ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras. *IEEE Transactions on Robotics*, **33**(5), 1255-1262, <https://doi.org/10.1109/TRO.2017.2705103>.
- RUSU, R. B. & COUSINS, S., 2011: 3D is here: Point Cloud Library (PCL). *Proceedings - IEEE International Conference on Robotics and Automation*, 1-4. <https://doi.org/10.1109/ICRA.2011.5980567>.
- STURM, J., ENGELHARD, N., ENDRES, F., BURGARD, W. & CREMERS, D. 2012: A benchmark for the evaluation of RGB-D SLAM systems. *IEEE International Conference on Intelligent Robots and Systems*, 573-580, <https://doi.org/10.1109/IROS.2012.6385773>.
- ZHANG, H., XU, C. & GU, J., 2022: Dense Reconstruction from Visual SLAM with Probabilistic Multi-Sequence Merging. *Canadian Conference on Electrical and Computer Engineering*, 2022-September, 33-40, <https://doi.org/10.1109/CCECE49351.2022.9918256>.

# Mobile 3D-Vermessung zur Dokumentation von Gasleitungen

THEA MINNICH<sup>1</sup> & ANSGAR BRUNN<sup>1</sup>

*Zusammenfassung: In den letzten Jahren haben sich tragbare Mobile Mapping Systems (Handheld Mobile Mapping System, HMS) als schnelle, umfangreiche und effiziente Aufnahmemethode erwiesen. Dieser Artikel befasst sich mit der Entwicklung und Untersuchung von HMS für einfache vermessungstechnische Arbeiten bei der Dokumentation von Hausanschlussleitungen der Sparte Gas. Ergänzend bieten Smartphones erste Möglichkeiten der Objektmessung (Consumer Mobile Mapping System, CMS).*

*Das Anwendungsszenario liegt in der Gasversorgung, welche als wesentliche Komponente der Energieversorgung einen entscheidenden Bestandteil der kritischen Infrastruktur in Deutschland ausmacht. Anforderungen an die Netzdokumentation und damit auch die Verfahren werden anhand von Regelwerken festgelegt und müssen bei der Auswahl beachtet werden. Die neuen bild- und laserbasierten Verfahren haben das Potential, konventionelle Messmethoden in der Vermessung des Gasleitungsnetzes abzulösen. Durch die frühe Digitalisierung der Erfassung direkt an der Baugrube kann die Leitungserfassung genauer und wirtschaftlicher gestaltet werden.*

## 1 Einleitung

Die weltweite Energieversorgung für Industrie, Gewerbe und Privathaushalte beruht zu einem großen Teil auf einer zuverlässigen Gasversorgung. Als wesentliche Komponente der Energieversorgung macht die Gasversorgung somit einen entscheidenden Bestandteil der kritischen Infrastruktur aus. Allein in Deutschland liegen mehr als 500.000 km Gasleitungen. Hausanschlussleitungen, die lokalen Leitungen bis zum Endverbraucher, nehmen einen wesentlichen Anteil der vorhandenen Leitungen ein und stehen auch für neue Bauvorhaben im Fokus.

Bei Tiefbaumaßnahmen geht von Gasleitungen ein hohes Gefährdungspotential aus, wenn unbeabsichtigt Beschädigungen herbeigeführt werden. Um Störungen in der Versorgung zu vermeiden, ist für jedes Tiefbauvorhaben eine Leitungsauskunft einzuholen. Eine umfassende und genaue Dokumentation der Leitungen ist daher unverzichtbar für sichere Arbeiten und um unbeabsichtigte Unfälle zu vermeiden.

Die in offener Bauweise erstellten Gasleitungen sind zunächst erstmal nur direkt nach der Montage sichtbar und dokumentierbar, bevor der Graben wieder verfüllt wird. Verlegungsverfahren mittels Erdrakete u. ä. liegen außerhalb der Betrachtungen dieses Artikels.

Aktuelle Vermessungsarbeiten in der Energieversorgung werden je nach Komplexität von ausgebildeten Vermessungsfachkräften durchgeführt. Dazu werden Vermessungsinstrumente wie Tachymeter und GNSS verwendet. Bei geringerer Komplexität der Maßnahme und kurzen Leitungslängen obliegt es dem geschulten Personal vor Ort mithilfe einfacher Aufnahmemethoden, beispielsweise einem Messband, manuelle Aufnahmeskizzen zur Dokumentation anzufertigen. Die

---

<sup>1</sup> Technische Hochschule Würzburg-Schweinfurt, Fakultät Kunststofftechnik und Vermessung, Röntgenring 8, D-97070 Würzburg, E-Mail: Thea.Minnich@study.thws.de, Ansgar.Brunn@thws.de

Aufnahmeskizzen beinhalten Maße zur nahegelegenen Topografie anhand denen GIS-Mitarbeitende die Leitungen im Innendienst konstruieren und in das GIS einpflegen. Die Bemaßung aller Leitungen und Einbauteile anhand der Topographie, wie etwa Bauwerke, ist zudem Bestandteil der Planwerke, welche ein schnelles Wiederauffinden im Störfall ermöglichen.

Das Ziel dieses Beitrags ist die Entwicklung und Untersuchung von Messverfahren für diese einfachen Messsituationen der Gasleitungsdokumentation, in denen mit einfachen Messmitteln die Dokumentationsnotwendigkeit erreicht wird.

In den letzten Jahren haben sich tragbare Mobile Mapping Systeme (Handheld Mobile Mapping System, HMS) als schnelle, umfangreiche und effiziente Aufnahmemethode erwiesen. Ergänzend bieten Mobiltelefongeräte erste Möglichkeiten der Objektmessung (Consumer Mobile Mapping System, CMS).

## 2 Mobile und handgetragene LiDAR Systeme

Laserscanning ist sehr vielseitig und die Sensoren werden in verschiedenen Systemen und Konfigurationen verbaut. Vor allem mit Airborne Laserscannern und mobilen Mapping-Systemen können in geringer Zeit großflächige Kartierungen durchgeführt werden (ELHASHASH et al. 2022). Als Mobile Mapping-Systeme (MMS) werden eine Art von Laserscanning-Systemen bezeichnet, die auf einem Fahrzeug, z. B. einem Auto oder einem Boot, montiert sind oder von Personen als Rucksack oder in der Hand getragen werden. Sie dienen zur Erfassung von 3D-Daten der Umgebung, während das Fahrzeug bzw. die Person in Bewegung ist. HMS sind oft so konzipiert, dass diese bequem in der Hand gehalten werden können, tragbar sind und wenig Gewicht haben, so dass sie bei der Arbeit vor Ort und bei der Datenerfassung in engen oder schwer zugänglichen Bereichen leicht eingesetzt werden können.

Für die Positionsbestimmung werden MMS durch verschiedene Sensoren ausgestattet. Durch GNSS-Receiver, IMU, Odometer, Laserscanner, Kameras und verschiedenen Kombinationsmöglichkeiten der Sensoren kann die Position des Geräts und zugleich die Situation während der Messung bestimmt werden. In vielen HMS wird Simultaneous Localization and Mapping (SLAM) angewendet. SLAM-Algorithmen stammen aus der Robotik und Computer Vision und werden in mobilen Kartierungssystemen verwendet, um die Position und Ausrichtung der Sensoren sowie die Position und Ausrichtung des Laserscanners zu bestimmen. Der Einsatz von SLAM ermöglicht es dem System, selbst in Umgebungen ohne GNSS oder in dynamischen Umgebungen genaue 3D-Karten der Umgebung zu erstellen. Die Kombination von HMS mit verschiedenen Sensoren zur Positionsbestimmung und SLAM ermöglicht es dem System, 3D-Daten während der Bewegung zu erfassen, gleichzeitig seinen eigenen Standort zu bestimmen und eine Karte seiner Umgebung in Echtzeit zu erstellen, was zu äußerst genauen und detaillierten 3D-Modellen der Umgebung führt. Auch Robotic Systeme (RMS) müssen zukünftig berücksichtigt werden (MASET et al. 2021). Eine neue Entwicklung –seit 2020– von HMS ist der Einsatz von LiDAR-Sensoren in Smartphones. Diese werden als Consumer Mobile Mapping Systeme (CMS) bezeichnet. CMS sind Kartierungs- und Navigationssysteme, die für die Nutzung durch einzelne Verbraucher und ursprünglich nicht für professionelle Vermessungs- oder Geodatenanwendungen konzipiert wurden. Sie sind in der Regel in mobile Geräte wie Smartphones und Tablets integriert und nutzen ebenfalls GNSS

und andere Sensoren. Diese Systeme bieten den Nutzern Karten und standortbezogene Informationen, aktuelle Verkehrs- und Navigationsinformationen in Echtzeit sowie die Möglichkeit, nach interessanten Orten wie Restaurants, Tankstellen und Geschäften zu suchen. Der Detaillierungsgrad, die Genauigkeit und die Präzision dieser Systeme ist in der Regel geringer als bei professionellen HMS.

### **3 Netzdokumentation**

Um die Versorgungssicherheit und Qualität der Gasversorgung in Deutschland zu gewährleisten, gibt es verbindliche Regelwerke, die von Fachleuten wie Vermessungsingenieuren und -technikern in der Netzdokumentation beachtet werden müssen. Das Gasnetz wird grundsätzlich nach den Druckstufen Hochdruck, Mitteldruck und Niederdruck unterschieden. Bei den hier behandelten Hausanschlussleitungen wird ein Druck von maximal 100 mbar, Niederdruck, geführt. Als Hausanschlussleitung wird die Gasleitung bezeichnet, die an das öffentliche Gasnetz angeschlossen ist und von dort aus in das Gebäude führt. Meist ist diese mit einem Absperrventil versehen. Der Verantwortungsbereich für die Leitung liegt bis zu der Hausmauer bei dem zuständigen Netzdienstleister/Gasversorger. Der Netzdienstleister trägt die Verantwortung, die Netzdokumentation für alle Sparten zu führen und über die Lage und Tiefe der verlegten Leitungen bei Bedarf eine Auskunft zu erteilen.

#### **3.1 Einfache vermessungstechnische Aufnahmen**

Baustellen - in offener Bauweise - für Hausanschlüsse sind oft sehr übersichtlich und haben einen begrenzten räumlichen Umfang. Die Leitungsdokumentation kann damit als einfache wiederkehrende vermessungstechnische Arbeit durch, nach Regelwerk „GW128“ des Deutschen Vereins des Gas- und Wasserfaches (DVGW), geschultes Personal für die Einpflegung in das GIS aufgenommen werden. Anhand des Regelwerks GW 120 werden Voraussetzungen für einfache vermessungstechnische Arbeiten für die Leitungsdokumentation wie folgt definiert:

- Gebäude mit einfacher Grundrissgeometrie,
- Möglichkeit auf nahe gelegene, in den Geobasisdaten vorhandene Gebäude oder sonstige geeignete topografische Objekte einzumessen,
- eine weitestgehend ebene Topografie,
- einfache Leitungsverläufe.
- Das längste gemessene Einzelmaß sollte bei der Verwendung eines Maßbandes i. d. R. nicht länger als 20 m sein (DVGW 2011).

Dazu können einfache Aufnahmemethoden, beispielsweise ein Messband, verwendet werden. Für jede dieser Baumaßnahmen muss eine Aufnahmeskizze für die Dokumentation erfolgen, die Maße zur nahegelegenen Topografie beinhaltet. Die Skizze kann manuell und nicht-maßstäblich erfolgen. Inhalte der Aufnahmeskizze sind Angaben über, die Gebäude, Straßennamen und Hausnummer, die Art der Leitung, Dimension, Material, Verbindungsart und Rohrumhüllung, Baujahr, die Deckung, Einbauteile, die Dimension und Material von Mantelrohren, die Anlagen, Reparaturstellen, Kabel, kathodische Korrosionsschutzanlagen sowie die ausführende Firma und Person sowie das

Datum (DVGW 2021). GIS-Mitarbeitende pflegen die Informationen aus der Aufnahmeskizze anschließend in den digitalen Bestandsplan ein.

### **3.2 Planwerke**

Netzpläne geben den Leitungsbestand wieder und dienen dem Nachweis der Lage von Leitungen. Planwerke werden in der Netzdokumentation unterschieden in Aufnahmeskizzen, Bestandspläne und Übersichtspläne. Übersichtspläne stellen das Versorgungsgebiet dar und beinhalten Haupt- und Versorgungsleitungen. Hausanschlüsse werden in den Bestandsplänen geführt. Grundlage für die Bestandsplanwerke, wie Flurstücksgrenzen, Polygonpunkte und Gebäude, entstammen den amtlichen Geobasisdaten. Digitale Datenmodelle zu den Leitungen werden im GIS mit Informationen nach Anhang A der GW 120 geführt.

## **4 Anforderungen**

Bei der Vermessung von Gasleitungen müssen einige Anforderungen berücksichtigt werden, um eine genaue und sichere Aufnahme zu gewährleisten. Dafür müssen der Einsatzbereich und die Genauigkeitsanforderungen abgegrenzt werden.

### **4.1 Spezifikation des Einsatzbereichs**

Der Einsatz soll zunächst im Rohrnetz, spezifischer in der Sparte Gas, anhand von neu gebauten Hausanschlüssen getestet werden. Für einen HA eines Einfamilienhauses werden im Standardfall Rohre aus Polyethylen (PE) mit einem Außendurchmesser von 32 Millimetern verbaut. Das Erfassungsgebiet liegt ausschließlich im Außenbereich. Reguläre HA liegen in einer offenen Baugrube mit begrenztem räumlicher Umfang. Als Referenzobjekt für die Positionierung sind Häuser aus den amtlichen Geobasisdaten denkbar oder mit konventionellen Vermessungsmethoden eingemessene Hausecken.

Weitere mögliche Einsatzbereiche für die Aufnahme mit LiDAR-Sensorik sind sanierte bzw. erneuerte Leitungen, beispielsweise nach einer Leckage, oder die Trennung / Außerbetriebnahme vorhandener Leitungen bei Baumaßnahmen oder Abriss. Außerdem ist auch eine Erweiterung des Einsatzgebietes auf Hauptleitungen, Versorgungsleitungen und Anschlussleitungen denkbar, welche einen weiteren räumlichen Umfang haben.

### **4.2 Technische Rahmenbedingungen und Bewertungskriterien**

Für eine mögliche Einführung der Systeme bei Netzdienstleistern müssen technische Rahmenbedingungen erfüllt werden. Die Wiederherstellung von Lage und Höhe der Leitungen darf mit einer maximalen Abweichung von 0,2 m erfolgen. Für Werte im  $3\sigma$ -Intervall muss eine Messgenauigkeit bei der vermessungstechnischen Aufnahme von 0,07 m eingehalten werden.

Um eine Aktualität der Daten zu gewährleisten, ist- neben dem Gesamtprozess der Netzdokumentation- die Dauer einzelner Teilprozesse nicht zu vernachlässigen. Somit ist auch die Dauer der Messung ein Bewertungskriterium. Dafür sollen die Performanz, die Schnittstellen und der Datenfluss der eingesetzten Messmethoden mit denen der üblichen Messmethoden verglichen werden. Weitere Bewertungskriterien sind die erreichbare Genauigkeit, die Reichweite und das Auflösungsanforderungen des LiDAR-Sensors. Weiter sollen der Funktionsumfang, die Handhabung

der Geräte und Applikationen, der Export und die Methoden der Georeferenzierung, beispielsweise über vorhandene Geobasis-Daten von Häusern evaluiert werden.

## **5 Marktanalyse HMS und CMS**

In den letzten Jahren haben sich HMS als effiziente Aufnahmemethode herausgestellt. Eine Vielzahl von Herstellern bieten mobile HMS an, welche unterschiedliche Spezifikationen und Funktionen aufweisen. Eine Auswahl wird hier aufgeführt.

### **5.1 HMS am Markt**

Eine Vielzahl von Herstellern bieten mobile HMS an, welche unterschiedliche Spezifikationen und Funktionen aufweisen (Tabelle 1). Dies beinhaltet verschiedene Typen von Sensoren, darunter LiDAR, Photogrammetrie und IMU, sowie Datenverarbeitungssoftware und Hardware. Einige dieser Anbieter stellen auch anwendungsspezifische Lösungen bereit, wie zum Beispiel industrielle Inspektionen, Vermessungen von Gebäuden und Infrastrukturen. Die hier aufgeführten HMS basieren alle auf unternehmensspezifischen SLAM-Algorithmen.

### **5.2 CMS am Markt**

CMS sind Kartierungs- und Navigationssysteme, die für die Nutzung durch einzelne Verbraucher und ursprünglich nicht für professionelle Vermessungs- oder Geodatenanwendungen konzipiert wurden. Sie sind in der Regel in mobile Geräte wie Smartphones und Tablets integriert und nutzen ebenfalls GNSS und andere Sensoren, um Navigations- und Kartierungsdienste anzubieten. Diese Systeme bieten den Nutzern Karten und standortbezogene Informationen, aktuelle Verkehrs- und Navigationsinformationen in Echtzeit sowie die Möglichkeit, nach interessanten Orten wie Restaurants, Tankstellen und Geschäften zu suchen. Der Detaillierungsgrad, die Genauigkeit und die Präzision dieser Systeme ist in der Regel geringer als bei professionellen HMS. Die Integration eines LiDAR-Sensors in Tablets oder Mobiltelefone bietet, neben ausschließlich vermessungstechnisch verwendeten HMS, vielversprechende Perspektiven für die Durchführung schneller Vermessungen.

#### **5.2.1 Apple iPhone 12 Pro**

Das iPhone 12 Pro von Apple wird, aufgrund seiner Verfügbarkeit zu einem geringen Preis im Vergleich zu traditionellen Vermessungsinstrumenten, seiner Tragbarkeit und der geringen Zeit, die für die Datenerfassung und -verarbeitung erforderlich ist, in diesem Kontext als alternative Option betrachtet.

Tab. 1: Herstellerangaben HMS

Marke	GeoSLAM (GEO-SLAM 2023a)	Leica (LEICA 2023)	GreenValley International (GREEN-VALLEY 2023)
Gerätename	ZEB Horizon	BLK2GO	LiGrip
Aussehen	 Abb. 1: ZEB Horizon (GEO-SLAM 2023b)	 Abb. 2: BLK2GO (LEICA 2023)	 Abb. 3: LiGrip (COMMUNICATIE FIZZ MARKETING 2023)
Anwendungsbereich	Innen- und Außenbereich	Innenbereich	Innen- und Außenbereich
Distanzbereich	Bis zu 100 m	0,5 bis 25 m	Bis zu 120 m
Gewicht mit Akku	3,7 kg	775 g	1,74 kg
Datenmenge	300.000 pts/sec Rohdaten: 15-50 MB/min	420.000 pts/sec	320.000 pts/sec
Software	- Post processing - Rohdaten: 15-50 MB/min - GeoSLAM Beam	- App für iOS und Android (live während des Scannens) - Grandslam	- eigener SLAM-Algorithmus
Relative Genauigkeiten [cm]	1-3 cm (bis zu 0,6 mm bei optimalen Bedingungen)	1 cm (indoor)	3 cm

Seit 2020 mit dem Release des Tablett iPad Pro 2020 und dem Smartphone iPhone 12 Pro verbaut Apple einen Light Detection and Ranging (LiDAR) Sensor in Smartphones und Tablets (APPLE 2023). Die von Apple vorgesehenen Funktionen zu der Unterstützung des Autofokus und Augmented Reality Anwendungen ermöglichen es jedoch nicht, eine Punktwolke als solche aufzunehmen. Hierfür bieten verschiedene Drittanbieter kostenpflichtige Applikationen an. Laut Herstellerangaben verfügt das iPhone 12 Pro über einen Beschleunigungssensor und ein 3-Achsen Gyrometer. Während der LiDAR-Sensor die Entfernung zu Objekten misst, können Gyrometer und Beschleunigungssensor verwendet werden und die Position und Ausrichtung der gemessenen Punkte in Bezug auf das Gerät zu bestimmen. Technische Daten des LiDAR-Sensors, wie die Auflösung, werden vom Hersteller nicht veröffentlicht (APPLE 2023).

### 5.2.2 Applikation „SiteScape“

Die Applikation „SiteScape“ generiert Punktwolken aus den Originaldaten des LiDAR-Sensors. In der App können Einstellungsmöglichkeiten für die Punktdichte und die Darstellung der Punktgröße unterschieden werden in „Low“, „Medium“ und „High“. „SiteScape“ gibt dabei nicht an, welche Auflösungen mit den jeweiligen Einstellungen erreicht werden können (SITESCAPE 2023).

### 5.2.3 Applikation „PolyCam“

PolyCam ist eine App für Fotografie und Videografie. Die App ermöglicht es Benutzern, mehrere Fotos aufzunehmen und diese in Echtzeit zu einem einzigen Panoramabild zu verarbeiten. Darüber hinaus bietet PolyCam auch Funktionen, um die Bildqualität zu verbessern. Polycam nutzt den Lidar-Sensor des iPhone, um eine räumliche Wahrnehmung und eine hochpräzise Tiefenerfassung von Objekten und Umgebungen zu ermöglichen. Dies kann bei verschiedenen Anwendungen, wie z.B. bei der Erstellung von 3D-Modellen, fotorealistischen Renderings und der Erfassung von Tiefen- und Entfernungsdaten, hilfreich sein. Für 3D-Modelle werden die LiDAR-Daten für die Erstellung von Punktwolken mit Hilfe der Photogrammetrischen Aufnahme gestützt (POLYCAM 2023).

## 6 Praxistest des iPhone 12 Pro LiDAR-Sensors

Genauigkeiten von mobilen Mapping Geräten (HMS und CMS) werden von Herstellern oft generisch angegeben. Im Fall des iPhone 12 Pro werden sogar gar keine Angaben zur Genauigkeit durch Apple Inc. veröffentlicht. Um die Genauigkeiten bei unterschiedlichen Rahmenbedingungen festzustellen, muss evaluiert werden, wie sich das System in expliziten realen Einsatzszenarien verhält. Bei Untersuchungen der Genauigkeit des iPad Pro 11s LiDAR-Sensors wurden die distanzabhängigen Auflösungen (s. Tab. 2) ermittelt. Dabei wurde ebenfalls die App „SiteScape“ verwendet (SPREFICA et al. 2021).

Tab. 2: Horizontale/vertikale Punktabstände

Einstellung der Punktdichte	Entfernung vom Objekt (m)			
	1	2	3	4
	Punktabstände / Auflösung (cm)			
High	0,9	1,9	2,6	3,3
Medium	1,3	2,6	3,5	4,6
Low	1,8	3,6	4,9	6,6

Eine weitere Bewertung stellt fest, dass es keinen Unterschied zwischen den LiDAR-Sensoren des iPhones und des iPads gibt. Weiterhin sei eine absolute Genauigkeit von einem Zentimeter für Formen kleiner Objekte und ein Präzisionsfehler von einem Zentimeter gemessen worden. Die Präzision nehme ab, wenn Oberflächen mit einer Seitenlänge von weniger als 10 cm gescannt



werden, und die Grenze für die Erkennung von Objekten liege bei etwa fünf Zentimetern (LUETZENBURG et al. 2021).

## 6.1 Teststellung

Bei dem Test sollen die Genauigkeiten und Auflösung der entstandenen Punktwolken und Abweichungen in einem realen Szenario im Außenbereich untersucht werden. Die Testmessung erfolgt an einem neu verlegten Hausanschluss (Abb. 4). Analog erfolgt eine tachymetrische Aufnahme, mit freier Stationierung über GNSS, der Hausanschlussleitung, sowie der Gebäudeaußenseite, mit der die Genauigkeit der Punktwolke verglichen werden soll.

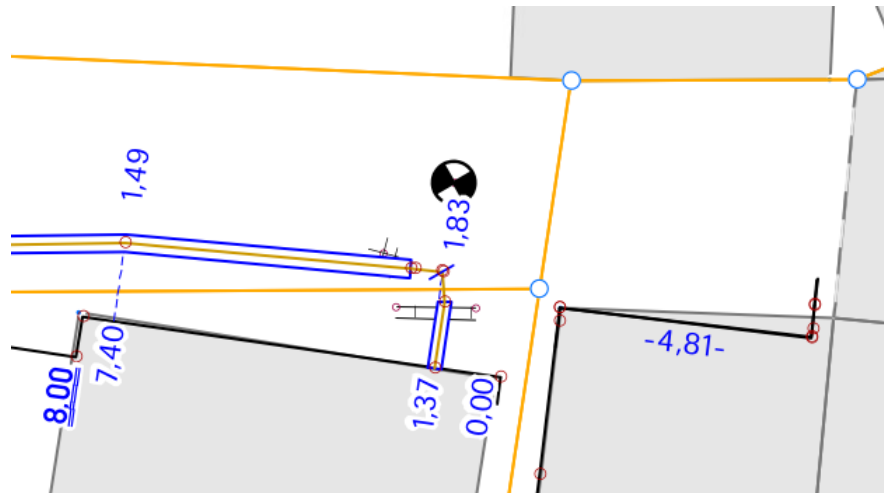


Abb. 4: Aufnahmeskizze vor Ort

## 6.2 Testmessung

Mehrere Messungen erfolgen mit der App „SiteScape“. Für alle Messungen wurde die Einstellung „High“ für die Punktdichte verwendet. Bei den Messungen wurde ein Abstand von ca. 1-2 Metern zwischen dem Smartphone und der Leitung gehalten. Der Abstand zu den Hausaußenwänden betrug 2-3 Meter.

Die erste Messsituation mit „SiteScape“ erfolgt mit geneigtem Smartphone über der Baugrube und schließt außerdem die Messung, senkrecht zu den Hausaußenwänden, von zwei Hausseiten ein. Es wurde auf eine gleichmäßige und langsame Bewegung des Smartphones und Schleifenschlüsse geachtet, sodass die Baugrube einmal während der Aufnahme umrundet wurde. Die Messung dauert 2-3 Minuten.

Bei einer zweiten Messung wurde das Smartphone, bei sonst gleichbleibenden Bedingungen, horizontal über der Baugrube bewegt und nicht geneigt. Bei einer dritten Messsituation wurde die Baugrube (Abb. 6) betreten und die Leitung aus einer Entfernung von ca. 0,6 m einseitig aufgenommen.

Zusätzlich wurde mit der Applikation „Polycam“ eine Probeaufnahme gemacht, um die Verknüpfung mit photogrammetrischen Ansätzen als Stütze für die LiDAR-Daten zu testen. Die Messsituation erfolgt erneut mit geneigtem Smartphone über der Baustelle und wurde von oben in die Baugruppe gehalten.

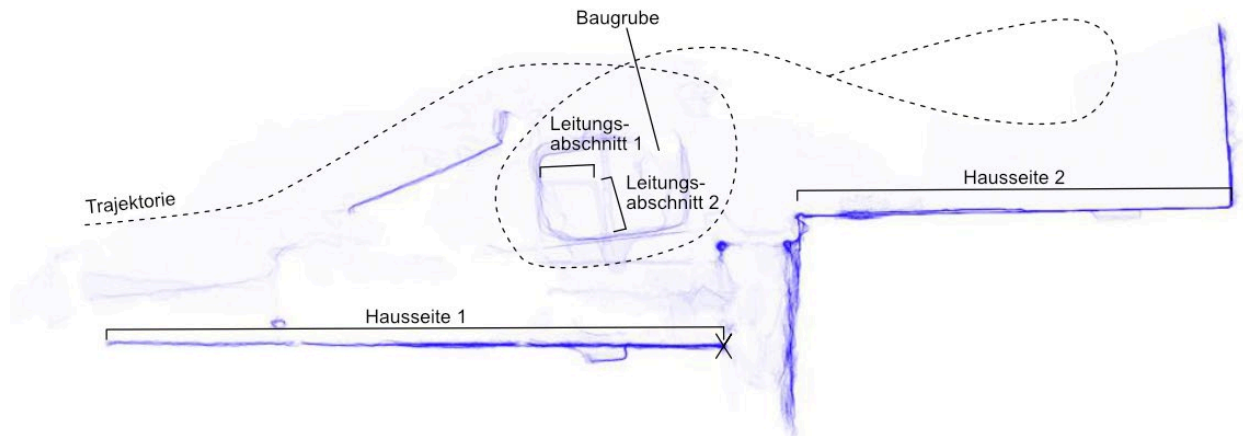


Abb. 5: Übersicht Baustelle



Abb. 6: Baugrube

### 6.3 Ergebnisse und Bewertung

Die Applikationen „SiteScape“ und „Polycam“ sind beide sehr intuitiv bedienbar. Das ist teils darauf zurückzuführen, dass es wenige individuelle Einstellungsmöglichkeiten gibt. Die Aufnahmemethode ist sehr schnell und die Arbeitssicherheit kann erhöht werden, wenn die Baugrube nicht betreten werden muss. Durch die in Echtzeit entstehende Anzeige der gemessenen Punkte entsteht eine gute Übersicht des aufgenommenen Gebiets.

Die Verarbeitung der Punktwolke ist durch die Möglichkeit verschiedene Dateiformate zu exportieren bzw. in die Portale der Applikationen hochzuladen über open-source Produkte möglich, jedoch aufwändiger im Vergleich mit bisherigen manuellen Skizzen. Die Dateigrößen von 90 MB (Polycam) bis 180 MB (SiteScape) sind kleiner als zunächst erwartet und stellen in der Datenspeicherung keine Herausforderung dar.

Ein großer Nachteil des Smartphones gegenüber anderen HMS ist, dass keine Kalibrierung oder Prüfung des Sensors möglich ist. Reine LiDAR-Daten des iPhones ohne photogrammetrische Unterstützung eignen sich noch nicht, um eine Aufnahmeskizze in ausreichender Qualität herzustellen.

„Sitescape“ liefert auch in der detailliertesten Messkonfiguration keine ausreichende Qualität der Punktwolke, um daraus eine Aufnahmeskizze ableiten zu können. Mit einem geringen Abstand ist die Qualität der Messdaten zwar deutlich besser (s. Abb. 8), jedoch immer noch nicht ausreichend, um die Lage, Länge oder weitere Attribute der Leitung zu bestimmen. Es kann von einem abstandsbezogenem Fehler ausgegangen werden. Die erwartete Genauigkeit von unter einem Zentimeter bei einem Abstand von ca. 0,6 Metern zum Objekt (Tabelle 2) wurde bei der Leitung nicht erfüllt.

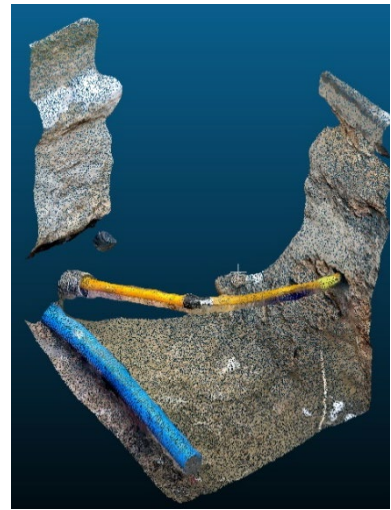
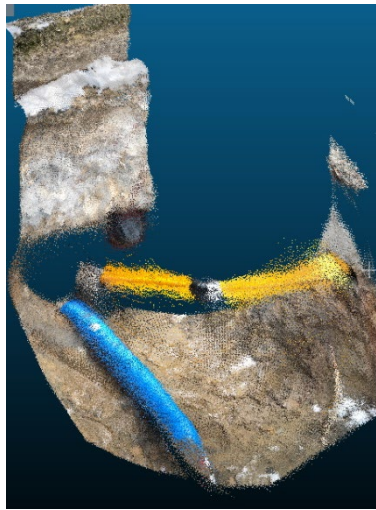
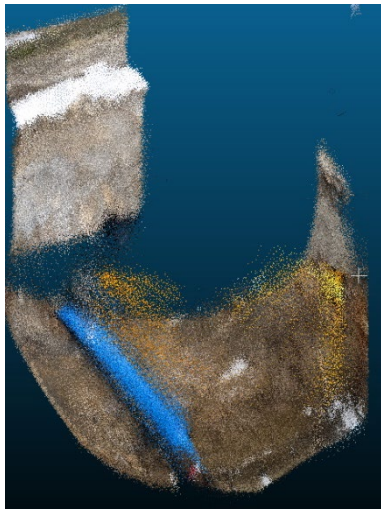


Abb. 7: Azimutale Aufnahme mit SiteScape

Abb. 8: Aufnahme in der Baugrube mit SiteScape

Abb. 9: Aufnahme mit Polycam

Auffällig ist bei den Messdaten, dass die Hausaußenseite als rechtwinklig aufgenommene Ebene sehr viel weniger Streuung aufwirft als die Leitung selbst. Die Referenzierung über die Hauskanten ist dahingehend möglich und wurde anhand der Messdaten aus „Polycam“ geprüft.

Eine Einbindung des LiDAR-Sensors des iPhones in einfache vermessungstechnische Aufgaben ist somit nur über weitere Sensorik denkbar. Anhand der – aus der Applikation „Polycam“ generierten – Messdaten (Abb. 9) kann die Länge der Leitung als auch die Deckung mit einer Abweichung von unter  $\pm 0,1$  m bestimmt werden (Tabelle 3). Die Koordinaten können nach einer Referenzierung über eine Hausaußenseite näherungsweise bestimmt werden. Die Abweichungen in der Lage von den Koordinaten der Punktwolke zu den Ist-Koordinaten betragen bis zu 20 cm. Das weitere Haus bzw. die Hausseite 1 wurde bei der Messung nicht vollständig erfasst und konnte daher nur annäherungsweise für die Referenzierung der Punktwolke verwendet werden, was die hohe Abweichung erklärt. Mit Abweichungen bis zu 9 mm in der Länge der Leitung und 2 cm in der Deckung werden die Genauigkeitsanforderungen eingehalten und es kann in diesem Fall kein Maßstabsfehler festgestellt werden.

Tab. 2: Zusammenstellung der Messergebnisse

	Soll-Länge [m]	Ist-Länge (Punktwolke) [m]	Abw. Länge [m]	Soll-Deckung [m]	Ist-Deckung [m]	Abw. Deckung [m]
Leitungsabschnitt 1	0,616	0,607	-0,009	0,70	0,72	+0,02
Leitungsabschnitt 2	0,621	0,624	+0,003	0,81	0,82	+0,02
Hausseite 1 (SiteScape)	7,991	8,10	+0,10	-	-	-
Hausseite 2	4,809	4,82	-0,01	-	-	-

## 7 Fazit

Die Tests an der Baugrube haben gezeigt, dass die Verwendung von CMS bei einfachen vermessungstechnischen Arbeiten der Gasleitungsdokumentation möglich ist und die Genauigkeitsanforderungen erfüllt werden können. Die Qualität von Punktwolken, die mit „SiteScape“ allein mit dem, im iPhone 12 Pro verbauten, LiDAR-Sensor reicht noch nicht für eine Objekterfassung von Leitungen aus. Ebenen, wie Häuser, dagegen sind mit einer ausreichenden Qualität erfasst worden. Mit einer Kombination von Sensorik bei „Polycam“, wie es auch bei professionellen HMS eingesetzt wird, ist eine zukünftige Verwendung für die vermessungstechnische Aufnahme von HA vorstellbar. Eine Übertragbarkeit auf weitere Netzinfrastruktur ist offensichtlich. Die Behandlung von komplexeren vermessungstechnischen Arbeiten mit CMS und professionellen HMS ist in weiteren Arbeiten notwendig.

## 8 Literaturverzeichnis

- APPLE INC., 2023: iPhone 12 Pro - Technische Daten (DE). [https://support.apple.com/kb/SP831?locale=de\\_DE](https://support.apple.com/kb/SP831?locale=de_DE), letzter Zugriff 06.02.2023.
- COMMUNICATIE FIZZ MARKETING, 2023: LiGrip. <https://geo-matching.com/handheld-laser-scanners/ligrip>, letzter Zugriff 06.02.2023.
- DVGW, 2011: DVGW GW 128 (M) Einfache vermessungstechnische Arbeiten an Versorgungsnetzen. Deutscher Verein des Gas- und Wasserfaches e.V. (HRSG.).
- DVGW, 2021: DVGW GW 120 Netzdokumentation in Versorgungsunternehmen. Deutscher Verein des Gas- und Wasserfaches e.V. (Hrsg.).
- ELHASHASH, M., ALBANWAN, H., QIN, R. 2022: A Review of Mobile Mapping Systems: From Sensors to Applications, Sensors, **22**(11), 4262.
- GEOSLAM, 2023a: Horizon Spec Sheet. [https://geoslam.com/wp-content/uploads/2021/03/Horizon\\_Spec\\_Sheet.pdf](https://geoslam.com/wp-content/uploads/2021/03/Horizon_Spec_Sheet.pdf), letzter Zugriff 06.02.2023.
- GEOSLAM, 2023b: ZEB Horizon: The Ultimate Mobile Mapping Solution. <https://geoslam.com/solutions/zeb-horizon>, letzter Zugriff 06.02.2023.

- GREENVALLEY INTERNATIONAL, 2023: LiGrip Specifications. <https://greenvalleyintl.com/static/upload/file/20220726/1658806022185090.pdf>, letzter Zugriff 06.02.2023.
- LEICA, 2023: Leica BLK2GO - Mobile Mapping. <https://shop.leica-geosystems.com/jp/ja-JP/leica-blk/blk2go>, letzter Zugriff 06.02.2023.
- LUETZENBURG, G., KROON, A. & BJØRK, A., 2021: Evaluation of the Apple iPhone 12 Pro LiDAR for an Application in Geosciences. *Scientific Reports*, **11**(1).
- MASET, E., SCALERA, L., BEINAT, A., CAZIRZI, F., CROSILLA, F., FUSIELLO, A. & GASPARETTO, A. 2021: Preliminary Comparison Between Handheld and Mobile Robotic Mapping Systems, *Proceedings of I4SDG Workshop 2021*, **108**.
- POLYCAM, 2023: Polycam Learn – FAQ. <https://learn.poly.cam/product-faqs>, letzter Zugriff 06.02.2023.
- SITESCAPE, 2023: FAQ – SiteScape. <https://www.sitescape.ai/faq>, letzter Zugriff 31.01.2023.
- SPREAFICO, A., CHIABRANDO, F., TEPPATI LOSÈ, L. & GIULIO TONOLO, F., 2021: The iPad Pro Built-in Lidar Sensor:3D Rapid Mapping Tests and Quality Assessment. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, **43**(B1-2021), 63-69.

# Real-time Distribution of an Airborne Situational Picture into Command and Control Systems

**DANIEL HEIN<sup>1</sup>; MATTHIAS GESSNER<sup>1</sup>, THOMAS KRAFT<sup>1</sup>, JULIA GONSCHOREK<sup>1</sup>,  
HENRY MEIßNER<sup>1</sup>, TILMAN BUCHER<sup>1</sup> & RALF BERGER<sup>1</sup>**

*Abstract: The Modular Aerial Camera System (MACS) has been developed, built and operated for more than a decade at the Institute of Optical Sensor Systems, German Aerospace Center (DLR, Berlin). It is a highly flexible system which is adapted to a wide range of carrier systems like Unmanned Aerial Systems (UAS), helicopters or piloted aircrafts. It is used for a variety of applications like mapping of environmental changes, 3D-reconstruction and urban mapping. One of the main goals of the system is to provide fast and reliable georeferenced information and situational awareness for civil security applications.*

*In this paper we present the most recent developments of MACS, enabling the integration of georeferenced image mosaics in real-time into command and control (C2) systems, GIS-software and mobile devices for first responders. The use of satellite communication systems allows the worldwide use of MACS even in destroyed environments without telecommunication services. The georeferenced image mosaics are disseminated to end users worldwide via web-map services. The developments are illustrated along several use cases including forest-fire and flooding.*

*The transfer of selected scientific developments and technologies to operational use and integration into C2 systems is done with commercial partners as part of the Helmholtz Innovation Lab OPTSAL. The workflow has successfully been certified to be integrated into a Web Map Service standard protocol, so the MACS-data can be shared in GIS systems worldwide. For disaster relief situations we demonstrated a workflow for integration and distribution of our live-map to all teams via the United Nations (UN) International Search And Rescue Advisory Group (INSARAG) coordination management system. Further developments include the use of onboard-classification to extract relevant information and reduce the amount of data to be transferred.*

## 1 Introduction

This article presents an operational workflow to produce an airborne situational picture and distribute it to a C2 system in real-time (RT). We give an overview about motivation (section 1), existing workflows and solutions with focus on search-and-rescue (SaR) and civil-security applications and their limitations (section 2). Then we outline our workflow and the technical prerequisites (section 3) and finally we illustrate the process along the example of situational pictures for first responders (section 4) and summarize the findings (section 5).

---

<sup>1</sup> German Aerospace Center (DLR), Institute of Optical Sensor Systems, Rutherfordstr. 2, D-12489 Berlin, E-Mail: [Daniel.Hein, Matthias.Gessner, Thomas.Kraft, Julia.Gonschorek, Henry.Meissner, Tilman.Bucher, Ralf.Berger]@dlr.de

## 1.1 Motivation

In times of climate change and increasing frequencies of weather extremes causing floods and forest fires even in northern and central Europe, the need for fast airborne situational pictures has emerged for coordinating emergency response more effectively. The successful management of major emergencies requires reliable and up-to-date information. A prominent example for such a situation was the flash flooding in the Ahr Valley (Germany) after heavy rainfall in July 2021. DLR was able to provide several large-scale operational pictures of affected regions within shortest time (DLR-OP 2021; WIELAND et al. 2022) to support the relief and rescue measures.

The Institute of Optical Sensor Systems (OS) at the German Aerospace Center (DLR) in Berlin is working to operationalize a novel technology for civil security applications to enable airborne situational pictures in real-time for first responders. DLR-OS is involved in an international SaR activity which is coordinated by the United Nations INSARAG working group. In cooperation with the non-profit aid organization International Search and Rescue Germany (I.S.A.R Germany) (DLR 2016) a first prototype of a MACS camera system was developed and tested to quickly retrieve airborne situational pictures by using fast flying drones (HEIN et al 2019). The further development of the prototype was done within the research projects Live-Lage (DLR 2018; BERGER et al. 2018) and INGENIOUS (INGENIOUS 2023).

Furthermore, DLR-OS operates the Helmholtz Innovation Lab OPTSAL (Optical Technologies for Situational Awareness) (OPTSAL 2023) with the goal to operationalize its camera technology and workflows together with industry partners, public authorities and end-users. In this context a close cooperation with EuroCommand has been established. The main product of EuroCommand is the C2 system CommandX (EUROCOMMAND 2022), which is broadly used by public authorities. Another cooperation was established with the drone supplier Quantum-Systems, which provides Vertical Take-Off and Landing (VTOL) drones (QUANTUM-SYSTEMS 2023) as carrier for MACS camera systems. Operational testing is being carried out in cooperation with the Duisburg Fire Department (FEUERWEHR DUISBURG, 2022).

## 1.2 Remote sensing situational pictures

Remote sensing based situational pictures are provided on different scales. On large scale satellite data are used. Optical satellites provide geocoded image maps and can be interpreted e.g. to provide flooded areas or forest fires when the view is not restricted by clouds, fog or smoke. Radar can be used in all weather conditions, yet the field of applications is restricted, a typical use case is e.g. flooding. The DLR Center for Satellite based Crisis Information (DLR-ZKI) in Oberpfaffenhofen (Germany) provides satellite based rapid situational image maps in cooperation with the Federal Office for Civil Protection (BBK) as part of the European wide Copernicus Emergency Management Service (CEMS). CEMS is operated 24/7 by a network of designated authorities, the National Focus points (COPERNICUS 2023a). Part of this is the Rapid Mapping service (COPERNICUS 2023b), which offers four standard products, a pre-event reference product and three post-event products: delineation, grading and first estimate. Typically, the products are provided within several hours to several days after the CEMS is activated. An overview over the service and typical activation times is given in (WANIA et al. 2021).

Airborne data has so far been limited in use in the CEMS, one example was the EMSR177 in 2016 covering an earthquake area in Central Italy one day after the event. Still the time from data acquisition to delivery has so far been considerably long. The airborne EMS will be extended by a manned aerial service operated by CGR/EuroSense and an unmanned aerial service operated by FairFleet. A 4-year framework contract has been awarded to these companies (EU 2022). The service will be provided where satellite images are not available or where the satellite's resolution is not sufficient.

Preparing fast-response airborne situational pictures has been a focus of DLR for several years. DLR has produced airborne image mosaics from manned aerial carriers e.g. for the Luebtheen forest fire 2019 (BUCHER et al. 2020) and the Ahr valley flash flooding (WIELAND et al. 2022, CRISIS PREVENTION 2021) within a few hours after data acquisition.

Drone based situational pictures have become widely available in the last years. Many public safety authorities and organizations (German acronym: BOS) operate specialized drone-teams (BOS DROHNENEINHEITEN DEUTSCHLAND, 2022). So far mainly live video streams are used to help the operational command to evaluate the situation. For drone based real-time mapping applications a small number of software solutions exist, e.g. DroneDeploy (DRONEDEPLOY 2022) and DJI Terra (DJI 2023), further details see section 2.

### **1.3 Real-time situational pictures in C2 systems**

C2 systems are used to provide a common situational picture to local authorities in case of an emergency situation and to distribute the information to the first responders. Maps and base layers stored on a server or locally can be displayed and shared with all instances connected to a Geographic Information System (GIS) or C2 system. In dynamically changing situations layers can be updated and added. In many cases this is done interactively by the first responders on mobile devices coordinated by the operational command. To add or update base layers and aerial imagery data sets the technical and geographic specifications and interfaces have to be met. So far, most C2 systems do not have the capabilities to generate aerial imagery base layers themselves within the software in real-time. Therefore pre-processed georeferenced data from third-party sources or software must be loaded, often as geo-tiffs. This adds a time-offset for map generation in third party tools and data transfer, which adds up to at least a couple of minutes after acquisition for small image maps, which are mostly drone based. Usually it takes hours to days for large areas until the data are processed and available.

## **2 Real-time data processing and distribution into C2 systems**

Real-time situational pictures from airborne imagery data rely on a) synchronous acquisition of image, position and attitude data, b) georectification and optimization of every image, c) continuous transmission of optimized image part to the ground, and d) transfer of the resulting image mosaic into a C2 system within a very short time. The georectification and image optimization process can be conducted in a separate software or in the C2 system itself, but it is preferable to do this instantaneously after image acquisition. To enable the real-time capability a transfer of the aerial image data is mandatory. This can be done by using several appropriate radio



links (i.e. WiFi, LTE, SatCom, StarLink, etc.). No matter what types of radio links are used the bandwidth is always a limiting factor and has to be considered. These four steps are shortly described in the following sections.

## 2.1 Acquisition of airborne image, position and attitude data

Georectification of airborne imagery data can be done in different ways. Most efficient and widespread is a parametric georectification based on the exact GNSS time, position (X, Y, Z) and attitude (Omega, Phi and Kappa) recorded while the image is taken. These parameters (exterior orientation) can be measured by using an appropriate GNSS aided Inertial Navigation System (INS). The exact synchronization between GNSS/INS system and the camera system needs to be ensured and all data must be stored in an appropriate manner. To gain good results the offset between GNSS antenna and IMU should be given as well as the offset between camera system and IMU. The internal geometry of the camera system (interior orientation) should be known as well to avoid lens distortions. The frame rate and the exposure time of the camera system can be a limiting factor in relation to the flight speed and altitude above ground. While high quality imaging is widespread for many UAV camera systems, quality of positional and attitude data can be limited or data even be not available for some of the smaller and cheaper systems. Access to all mentioned parameters is often restricted and leads to bad results in the georectification process.

## 2.2 Direct Georeferencing / Georectification

A small number of software solutions are available for drone based real-time mapping applications. DroneDeploy offers a live mapping feature with a reduced resolution by a factor of 5 (KASSIGKEIT 2021). It is so far of limited use for many public authorities due to data privacy conflicts as the data can only be transferred in the cloud. DJI Terra provides a near real-time feature with some observed latency (KASSIGKEIT 2021), which is allegedly caused by limited bandwidth of the UAS radio link. For none of these software solutions a live map export into C2 systems is provided. PIX4Dreact (Pix4D 2023) is a rapid mapping software which provides maps after landing. A delay of several minutes is present because of transfer of the raw data to a computing unit (i.e. laptop or work station) and the photogrammetric processing of the imagery on the ground. The computing speed of Pix4Dreact can be optimized by using modern graphics processing units (GPU). Pix4Dreact can handle input data of a variety of drones and camera types. The resulting map can be exported as GeoTiff.

A different way to geocode imagery is to use reference data and align the newly acquired data by selecting identical features or ground control points in both data sets. This is a time consuming offline-workflow, the results often are of limited quality, especially when strong local relief hampers the use of parametric models. Image and phase correlation have also been used to co-register image data to a base-map. An example for such a co-registration workflow for airborne and satellite data is COSI-COR (LEPRINCE et al. 2007).

Deep learning and AI have emerged as powerful tools for image co-registration. The software Edgybees (EDGYBEES 2023) rapidly co-registers satellite scenes in real-time to a reference map or fine-tunes and stabilizes coarsely geocoded video-livestreams e.g. from drones with limited orientation accuracy using a reference scene. Vice-versa reference data (vectors) can be re-

projected into the video-stream. This has already been applied in use cases with emergency responders (SWEET 2021).

Alternatively, a representation of an aerial image to a spatial grid can be calculated in consideration of a digital elevation model (DEM) and the mentioned interior and exterior orientation parameters. The earth's surface was measured by NASA with the Space Shuttle Radar Topography Mission (SRTM). The resulting DEM (SRTM 30) covers the whole earth and is available as open data set. Latest DEM data sets are available by TanDEM-X mission which is managed by DLR. TanDEM-X 90 is available as open data set.

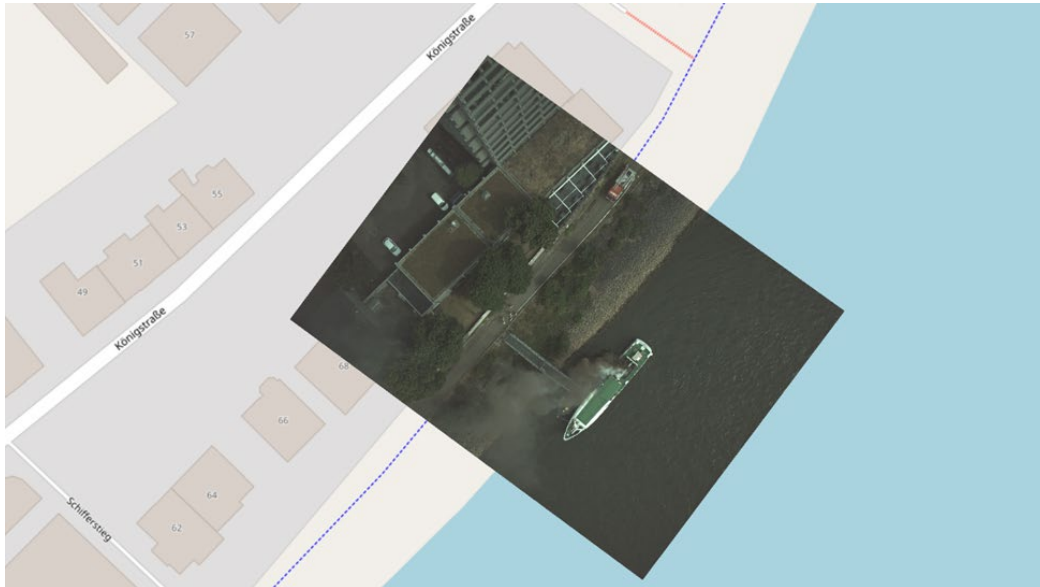


Fig. 1: Georectified aerial image of MACS using the SRTM digital elevation model

### 2.3 Transmission to the ground

The transfer of image and meta data from a drone to the ground is a restricting factor due to limited bandwidth, range and achievable data transmission rates during flight. Several radio transmission technologies are available like wireless local networks, telecom networks and satellite communication. The choice of the appropriate technology depends on the availability of the service and the technical requirements like size, weight, power, price, frequency range, bandwidth, data transfer rate and achievable range. No matter which technology is used, range and achievable data rate can strongly vary. To counteract these limitations the amount of data to be transferred should be minimized e.g. by reducing image size and image resolution. Furthermore, the data can be compressed. To avoid data loss buffering of data should be implemented as well.

### 2.4 Transfer of a georectified image data into a C2 system

Modern C2 systems have a GIS feature to manage, analyze and visualize geographic data. Most common is the import of GeoTiff files and/or using a Web Map Service (WMS).

It is recommended that the georeferencing and processing of the aerial image data is done outside of the C2 system to guarantee a fast import and quick visualization. This can be done by using individual or tiled files or by providing a WMS. Pre-processed files can be provided via USB stick,

hard disk or FTP server. A WMS needs additional infrastructure and maintenance yet it is recommended for real-time mapping applications despite the need for access to the internet and extra costs for maintenance and infrastructure.

### 3 Real-time situational picture by DLR

DLR has developed a suite of camera systems, the Modular Airborne Camera System MACS (LEHMANN et al. 2011; DLR 2023a) and software solutions (DLR 2023b) to operate these systems. In the last years the focus has moved to reducing the size of the systems (camera, computing stack, storage media) while maintaining and improving image quality and onboard processing capabilities. Another focus was to improve the real-time capabilities and diversify the options for data communication.

The general MACS camera concept has been described in (BRAUCHLE 2015). It combines a modular selection of calibrated cameras of variable spectral and geometric properties with high end GNSS/INS for accurate position and attitude data. A computing stack handles the onboard camera operation. A description of a marine airborne real-time situational picture acquired from a motorglider is given in BRAUCHLE et al. (2018). A description of an UAS mapping system including hard- and software is presented by HEIN et al. (2017). The processing was improved by using TAC (Terrain Aware Image Clipping) for direct georeferencing (HEIN 2018), which crops the images only retaining a small overlap with the neighboring images, thereby reducing the data to be downlinked up to a factor of 10 when high overlap e.g. 90% is recorded. The remaining clipped images are further compressed using a 12bit JPEG compression algorithm, which leads to a reduction of down to <1% of the original image size. This allows to retain the full geometric and radiometric resolution even with lower transfer rates. The whole workflow including the ground processing is described in detail in HEIN (2019).

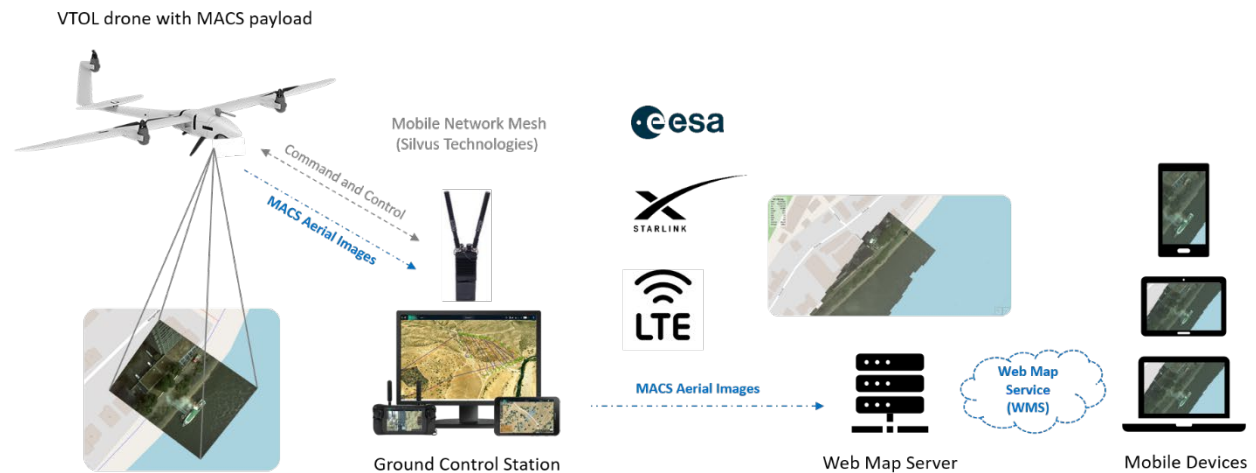


Fig. 2: Overview of the real-time mapping workflow with MACS and VTOL drone (Source: DLR)

The real-time mapping feature is available with the latest MACS development for the VTOL drone Vector by the German manufacturer Quantum Systems (QUANTUM SYSTEMS, 2022). As a result compressed georeferenced aerial image snippets are continuously received on the ground with one

or several distributed receivers. Incoming snippets are added to a live image mosaic which is continuously updated using the MACS RT-Viewer software (DLR, 2023b). The operator can navigate through the live map, select zoom stages and regions of interest. The histogram can be interactively stretched and adjusted to the operator’s needs. The software is able to export regions of interest as GeoTiff file. This allows sharing of data even if no telecommunication service is available. If a service is available (e.g. LTE, StarLink or IRIS<sup>2</sup>) the incoming aerial image snippets can be forwarded from the ground station to an appropriate Web Map Server. The resulting WMS can be consumed worldwide via GIS or C2 systems. This allows an easy integration of the live map even on mobile devices. The first prototype was tested and evaluated in 2021 by DLR in cooperation with several partners and institutions. The implementation of an appropriate WMS feature is still ongoing.



Fig. 3: VTOL drone Vector with latest MACS payload (left) and notebook with radio link module (right) to receive and visualize the situational picture (Source: DLR)

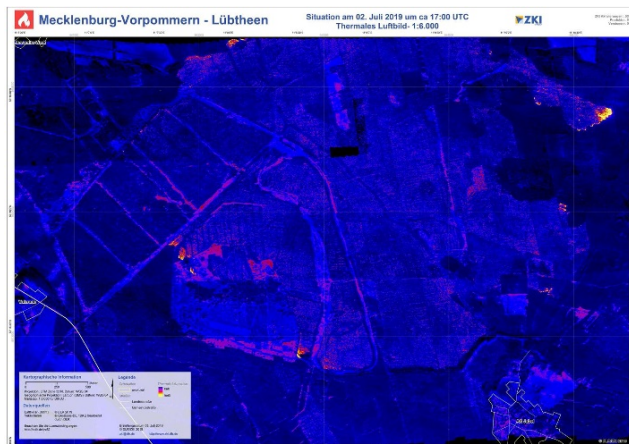


Fig. 4: Luebtheen forest fire situational map.



Fig. 5: Ahr Valley situational map.

## 4 Case studies

DLR has supported German authorities in several emergencies with situational pictures in the last years. In case of the Luebtheen forest fire (BUCHER 2019) or the Ahr Valley flash flood (2021) the image maps were taken on short notice within several hours with a manned airplane. The data were

not downlinked in real-time. It was transferred after landing via hard disk to DLR, processed and uploaded to the situational command via FTP server. Standardized image maps were processed by DLR ZKI (Fig. 4 and Fig. 5). Meanwhile a communication infrastructure has been established as further data communication link for MACS to distribute the live mapping result in real-time as Web Map Service.

In case of the experiments in the project LiveLage (DLR, 2022a) and the Grunewald forest fire (DLR, 2022b) the situational picture was live streamed via LTE in the C2 systems of the involved fire departments by using a WMS. In cooperation with I.S.A.R Germany the usage of StarLink was successfully tested. The three scenarios are described in the following sections.

#### 4.1 Exercise with Duisburg fire department

The project Live-Lage started in 2018 as cooperation between Duisburg fire department, I.S.A.R Germany and DLR. The aim was to develop a new technology to rapidly generate a large-scale situational picture by using fast flying drones (BERGER et al., 2018). The latest prototype was tested in May 2022 during an exercise. The VTOL drone Vector was activated from the Duisburg fire department building to monitor a fire on a ship in the Duisburg harbor. Image data was captured and processed onboard of the drone with the latest MACS camera system. The georectified image snippets were transferred to the ground continuously using the drone's radio link. The data was synchronously visualized at several places on the ground and as well in the operational control center. The situational picture was constantly received to cover the changing situation. A short video documentation of the exercise can be found at the official website (STADT DUISBURG, 2022).

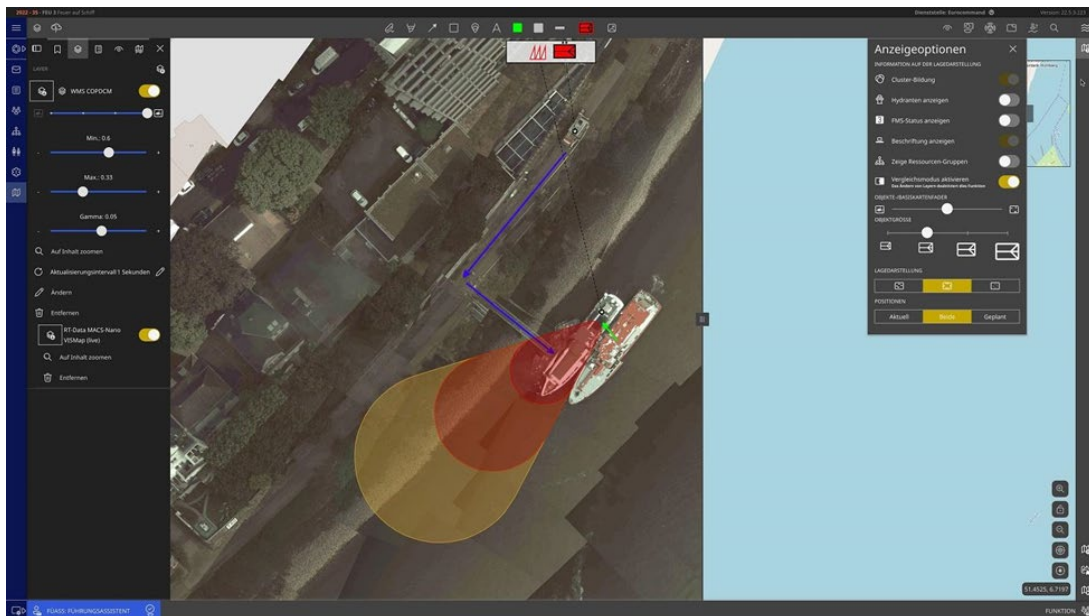


Fig. 6: RT-Situational picture at Duisburg Harbor in CommandX (image credit EuroCommand)

#### 4.2 Forest Fire Berlin-Grunewald

On August 3 2022 a fire on a site for controlled detonations in Berlin-Grunewald started, which led to uncontrolled detonations of ammunition. A security distance of 1000m was established, access to the fire was strongly reduced. Due to the fact that drones of the Berlin fire department

had a relatively small action radius due to the limited reach of their remote controllers, DLR was asked to provide a RT-situational map with its MACS-Nano system and a Vector VTOL drone, well suited for quickly mapping large areas (Fig. 7). The drone was started from the blocked AVUS motorway. A predefined flightpath was programmed and autonomously covered Beyond Visual Line of Sight (BVLoS). The georeferenced and compressed image snippets were transferred to the ground (BROADCAST SOLUTIONS, 2022) and forwarded to a server of the Berlin fire department, from which the data could be accessed in the C2 software in the mobile situational center (Fig. 8) and the mobile devices of the active firefighters.



Fig. 7: MACS-Nano in Vector UAS (Quantum systems)



Fig. 8: Visualization of MACS-data in a C2 system, Berlin fire department

### 4.3 Exercise with I.S.A.R Germany

In areas with destroyed infrastructure in disaster relief missions mobile satellite communication can be the key for transferring the data to decision makers. In exercises with I.S.A.R Germany we have repeatedly tested live-mapping for first responders. In May 2022 we successfully demonstrated the creation of a RT-situational map captured from a drone during flight and the transfer of the map via WMS to the United Nations INSARAG C2 System and in situation rooms worldwide via the StarLink satellite network of the US space company SpaceX (DLR 2022g).

## 5 Conclusions and Results

Airborne real-time situational pictures require a consistently performant and resource efficient workflow to handle a large amount of data while retaining radiometric and geometric quality and resolution. Delays caused by data redundancy can be solved by intelligent on-board processing like terrain aware clipping (image cropping) and compression, the risk of bottlenecks caused by bad downlink connections can be reduced by buffering and the selection of the most suitable communication links. Specific software for tone mapping and a real-time compatible viewer enable an on-the-fly adjustment correction and conversion to 8-Bit, which can be directly output to a WMS. This service can be provided using standard internet communication (LTE or StarLink) and it can be accessed synchronously worldwide via GIS, C2 software or simply a browser. This workflow was successfully employed in several real-world situations and exercises enabling access to a real-time situational picture to first responders. Work to further operationalize and

distribute these capabilities are ongoing in the Helmholtz Innovation Lab OPTSAL together with industry and BOS end-users.

## 6 Literaturverzeichnis

- BERGER, R., BAYER, S., HELMRICH, J. & KRAFT, T., 2018: Entwicklung und Erprobung eines neuartigen Konzeptes zur schnellen Aufklärung von Einsatzlagen - Eine Kooperation zwischen der Feuerwehr Duisburg und dem Deutschen Zentrum für Luft- und Raumfahrt (DLR). Jahresfachtagung der Vereinigung zur Förderung des Deutschen Brandschutzes e.V., 28.-30. Mai 2018, Duisburg. <https://elib.dlr.de/120806/>.
- BERLIN, 2022: Feuer nach Explosion im Grunewald. <https://www.berlin.de/aktuelles/7672047-958090-feuer-grunewald-explosion-verkehrseinsch.html>, last access 25.01.2023.
- BOS DROHNENEINHEITEN DEUTSCHLAND, 2022: Aufstellung über verschiedene BOS-Drohneneinheiten aus Deutschland. <https://drohneneinheit.de/einheiten/>, last access 25.01.2023.
- BRAUCHLE, J., HEIN, D. & BERGER, R., 2015: Detailed and highly accurate 3D models of high mountain areas by the MACS-Himalaya aerial camera platform. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., **40**(7/W3), 1129-1136, <https://doi.org/10.5194/isprsarchives-XL-7-W3-1129-2015>, 2015.
- BRAUCHLE, J., BAYER, S., HEIN, D., BERGER, R. & PLESS, S., 2018: MACS-Mar: a real-time remote sensing system for maritime security applications. CEAS Space Journal, 1-10, Springer, <https://doi.org/10.1007/s12567-018-0207-7>.
- BROADCAST SOLUTIONS, 2022: <https://broadcast-solutions.de/de/behoerdenkommunikation/produkte-loesungen/silvus-technologies/>, last access 25.01.2023.
- BUNDESAMT FÜR BEVÖLKERUNGSSCHUTZ UND KATASTROPHENHILFE, 2022: Drohnen im Bevölkerungsschutz. BBK, Ergebnisbericht zur Online-Befragung 2021, <https://www.bbk.bund.de/SharedDocs/Downloads/DE/Mediathek/Publikationen/Krisenmanagement/ergebnisbericht-online-befragung-2021-drohnen.pdf>, last access 25.01.2023.
- BUNDESREGIERUNG, 2020: Drohne dirigiert Rettungskräfte in Echtzeit. <https://www.bundesregierung.de/breg-de/service/archiv/sicherheitsforschung-live-lage-1806820>, last access 25.01.2023.
- CRISIS PREVENTION, 2021: DLR erfasst Schadenslage nach der Flutkatastrophe. <https://crisis-prevention.de/katastrophenschutz/dlr-er-fasst-scha-dens-la-ge-nach-der-flut-ka-ta-stro-phe.html>, last access 30.01.2023.
- COPERNICUS, 2023a: Copernicus Emergency Management Service – Mapping Services. <https://emergency.copernicus.eu/>, last access 25.01.2023.
- COPERNICUS, 2023b: Copernicus Emergency Management Service – Rapid Mapping. <https://emergency.copernicus.eu/mapping/ems/rapid-mapping-portfolio>, last access 25.01.2023.
- DJI, 2023: Product Website DJI TERRA. <https://www.dji.com/de/dji-terra>, last access 25.01.2023.
- DLR, 2016: DLR und I.S.A.R. - Gemeinsam für die schnelle Rettung. [https://www.dlr.de/content/de/artikel/news/2016/20161118\\_dlr-und-i-s-a-r-gemeinsam-fuer-die-schnelle-rettung\\_20120.html](https://www.dlr.de/content/de/artikel/news/2016/20161118_dlr-und-i-s-a-r-gemeinsam-fuer-die-schnelle-rettung_20120.html), last access 25.01.2023.

- DLR, 2018: DLR-Forschung hilft in Katastrophenfällen, [https://www.dlr.de/content/de/artikel/news/2018/2/20180530\\_dlr-forschung-hilft-in-katastrophenfaellen\\_27975.html](https://www.dlr.de/content/de/artikel/news/2018/2/20180530_dlr-forschung-hilft-in-katastrophenfaellen_27975.html), last access 25.01.2023.
- DLR, 2021: Lageinformationen für die Rettungskräfte. [https://www.dlr.de/content/de/artikel/news/2021/03/20210716\\_lageinformationen-fuer-die-rettungskraefte.html](https://www.dlr.de/content/de/artikel/news/2021/03/20210716_lageinformationen-fuer-die-rettungskraefte.html), last access 06.02.2023.
- DLR, 2022a: DLR und Feuerwehr erproben Lageaufklärung in Echtzeit. [https://www.dlr.de/content/de/artikel/news/2022/02/20220610\\_dlr-und-feuerwehr-erproben-lageaufklaerung-in-echtzeit.html](https://www.dlr.de/content/de/artikel/news/2022/02/20220610_dlr-und-feuerwehr-erproben-lageaufklaerung-in-echtzeit.html), last access 25.01.2023.
- DLR, 2022b: Präzise Luftbilddaten in Echtzeit für die Lagebeurteilung. [https://www.dlr.de/content/de/artikel/news/2022/03/20220812\\_praezise-luftbilddaten-in-echtzeit-fuer-die-lagebeurteilung.html](https://www.dlr.de/content/de/artikel/news/2022/03/20220812_praezise-luftbilddaten-in-echtzeit-fuer-die-lagebeurteilung.html), last access 25.01.2023.
- DLR, 2022c: Mit Luftbildkarten weltweit helfen. [https://www.dlr.de/content/en/articles/news/2022/04/20221205\\_providing-assistance-worldwide-with-aerial-maps.html](https://www.dlr.de/content/en/articles/news/2022/04/20221205_providing-assistance-worldwide-with-aerial-maps.html), last access 25.01.2023.
- DLR, 2022d: DLR und I.S.A.R. Germany verlängern Kooperationsvereinbarung. [https://www.dlr.de/content/de/artikel/news/2022/02/20220518\\_technologien-fuer-das-krisen-und-katastrophenmanagement.html](https://www.dlr.de/content/de/artikel/news/2022/02/20220518_technologien-fuer-das-krisen-und-katastrophenmanagement.html), last access 25.01.2023.
- DLR, 2022e: DLR Projekt Live-Lage. [https://www.dlr.de/os/desktopdefault.aspx/tabid-12893/22517\\_read-52089/](https://www.dlr.de/os/desktopdefault.aspx/tabid-12893/22517_read-52089/), last access 25.01.2023.
- DLR, 2022f: Lageinformation im Ahrtal. [https://www.dlr.de/content/de/artikel/news/2021/03/20210716\\_lageinformationen-fuer-die-rettungskraefte.html](https://www.dlr.de/content/de/artikel/news/2021/03/20210716_lageinformationen-fuer-die-rettungskraefte.html), last access 25.01.2023.
- DLR, 2022g: Providing assistance world-wide with aerial maps. [https://www.dlr.de/content/en/articles/news/2022/04/20221205\\_providing-assistance-worldwide-with-aerial-maps.html](https://www.dlr.de/content/en/articles/news/2022/04/20221205_providing-assistance-worldwide-with-aerial-maps.html).
- DLR, 2023a: Website Modular Aerial Camera System (MACS). <https://www.dlr.de/macs>, last access 25.01.2023.
- DLR, 2023b: MACS-Box - Software suite for the DLR developed Modular Aerial Camera System (MACS), <https://macs.dlr.de/box>, last access 25.01.2023.
- DRONEDEPLOY, 2022: Live Map. <https://www.dronedeploy.com/product/live-map/>, last access 25.01.2023.
- EDGYBEES, 2023: Product Website. <https://edgybees.com/>, last access 25.01.2023.
- EUROCOMMAND, 2022: Website der Eurocommand GmbH, <https://www.eurocommand.com/>, last access 25.01.2023.
- EU, UNION CIVIL PROTECTION KNOWLEDGE NETWORK, 2022: Enhanced capacity for the Copernicus Emergency Management Service. <https://civil-protection-knowledge-network.europa.eu/system/files/2022-10/Knowledge%20Network%20Newsletter%20-%20Issue%206%20-%20October%202022.pdf>, 18, last access 25.01.2023.
- FEUERWEHR DUISBURG, 2022: Das Forschungsprojekt Live-Lage. Manned and unmanned aerial services to complement satellite sources for humanitarian aid and civil protection. <https://www.duisburg.de/microsites/feuerwehr/feuerwehr/seh/livelage.php>, last access 25.01.2023.



- HEIN, D., BAYER, S., BERGER, R., KRAFT, T. & LESMEISTER, D., 2017: An Integrated Rapid Mapping System for Disaster Management. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, **42-1** (W1), 499-504, <https://doi.org/10.5194/isprs-archives-XLII-1-W1-499-2017>.
- HEIN, D. & BERGER, R., 2018: Terrain Aware Image Clipping for Real-Time Aerial Mapping. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, **IV-1**, 61-68, <https://doi.org/10.5194/isprs-annals-IV-1-61-2018>.
- HEIN, D., KRAFT, T., BRAUCHLE, J. & BERGER, R. 2019: Integrated UAV-Based Real-Time Mapping for Security Applications. *ISPRS Int. J. Geo-Inf.*, **8**(5), 1-16. <https://doi.org/10.3390/ijgi8050219>.
- HUNSDÖRFER, J., GONSCHOREK, J. & STEINHOFF, J., 2021: Mit Hightech aus Deutschland vor der Lage. *CP Crisis Prevention*, **3**(25), [https://www.flipsnack.com/betapublishing/cp\\_4-2021\\_inhalt-mit-umschlag/full-view.html](https://www.flipsnack.com/betapublishing/cp_4-2021_inhalt-mit-umschlag/full-view.html), last access 25.01.2023.
- INGENIOUS, 2023: Project Website. <https://ingenious-first-responders.eu/>, last access 31.01.2023.
- KASSIGKEIT, J., 2021: Vergleich von Verfahren zur Aufklärung zeitkritischer Einsatzlagen der Feuerwehr mithilfe georeferenzierter Drohnen-Luftbilder. Bachelorarbeit, Beuth University of Applied Sciences, <https://elib.dlr.de/143802/>.
- KRAFT, T., GEBNER, M., MEIBNER, H., CRAMER, M., GERKE, M. & PRZYBILLA, H. J., 2016: Evaluation of a metric camera system tailored for high precision UAV applications. *ISPRS Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, **41**(B1), 901-907, <https://doi.org/10.5194/isprs-archives-XLI-B1-901-2016>.
- LEHMANN, F. BERGER, R., BRAUCHLE, J., HEIN, D., MEISSNER, H., PLESS, S. STRACKENBROCK, B. & WIEDEN, A., 2011: MACS - Modular Airborne Camera System for Generating Photogrammetric High-Resolution Products. *Photogrammetrie - Fernerkundung – Geoinformation*, **6**, 435-446, <https://doi.org/10.1127/1432-8364/2011/0096>.
- LEPRINCE, S., BARBOT, S., AYOUB, F. & AVOUAC, J.P., 2007: Automatic and Precise Orthorectification, Coregistration, and Subpixel Correlation of Satellite Images, Application to Ground Deformation Measurements. *IEEE Trans. Geosci. Remote Sensing*, **45**(6), 1529-1558.
- OPTSAL, 2022: Helmholtz Innovation Lab OPTSAL (Optical Technologies for Situational Awareness), <https://www.optsal.de/>, last access 25.01.2023.
- PIX4D, 2022: Helping to protect the rainforest in Peru with drone mapping. <https://www.pix4d.com/blog/aerial-drone-mapping-amazon-rainforest/>, last access 25.01.2023.
- PIX4D, 2023: Pix4Dreact. <https://support.pix4d.com/hc/en-us/articles/360045923031-Supported-cameras-PIX4Dreact> ) last access 25.01.2023.
- QUANTUM SYSTEMS, 2022: Offizielle Website der Quantum Systems GmbH. <https://www.quantum-systems.com/>, last access 25.01.2023.
- STADT DUISBURG, 2022: Live-Lage Übung der Feuerwehr Duisburg am Binnenhafen. <https://youtu.be/7TfYeklolVc>, last access 25.01.2023.
- SILVUS TECHNOLOGIES (2022): Streamcaster Radios product description. <https://silvustechnologies.com/products/streamcaster-radios/>, last access 25.01.2023.

- SWEET, N., 2021: Edgybees Takes the Mystery out of Disaster Imagery. Inside Unmanned Systems, <https://insideunmannedsystems.com/edgybees-takes-the-mystery-out-of-disaster-imagery/>, last access 25.01.2023.
- WANIA, A., JOUBERT-BOITAT, I., DOTTORI, F., KALAS, M. & SALAMON, PM 2021: Increasing Timeliness of Satellite-Based Flood Mapping Using Early Warning Systems in the Copernicus Emergency Management Service. Remote Sensing, **13**(11), 2114, <https://doi.org/10.3390/rs13112114>.
- WIELAND, M., RESCH, B. & LECHNER, C., 2022: Entscheidungsunterstützung im Katastrophenschutz. CP Crisis Prevention, <https://crisis-prevention.de/kommunikation-it/entscheidungsunterstuetzung-im-katastrophenschutz.html>.

# Virtual Reality for Spatial Research in Archaeology: Prototyping 3D Reconstruction Tools for the Site of Yeha, Ethiopia

ALEXANDER P. WALMSLEY<sup>1</sup>, MAREN LINDSTAEDT<sup>1</sup>, MIKE SCHNELLE<sup>2</sup> & THOMAS P. KERSTEN<sup>1</sup>

*Abstract: There has been an abundance of research in recent years focusing on the potential of 3D reconstruction and visualisation technologies in cultural heritage and related fields. While many of these are focused on knowledge dissemination with a non-specialist audience, often within the context of museums or heritage sites, fewer studies have dedicated themselves to investigating how such technologies, including Virtual Reality, can be used as tools within the research process. While 3D virtual reconstruction technologies are often implemented by specialists of these technologies in consultation with the heritage researchers themselves, rarely are heritage researchers directly involved in the process of 3D modelling and visualisation. This article describes the development of a real-time 3D reconstruction toolset operated using a VR headset and motion controllers, in this case the HTC Vive. The case study used to test this tool is the archaeological site of Yeha, a town in the northern Tigray region of Ethiopia and the location of an archaeological site with monumental remains dating back to around 700 BCE. This site was chosen as archaeological excavations are currently under way and good 3D scanned data already exists. The 3D reconstruction tool allows archaeologists and other specialists to collaborate within VR to build a 3D visualisation of the ancient site that is able to be adapted as new information is brought to light.*

## 1 Introduction

3D virtual reconstruction is a powerful tool for both research and communication of cultural heritage. Advances in VR, AR, and web technologies over the past decade have allowed virtual 3D reconstructions to become increasingly widespread not only in the entertainment and tourism industries, but also in heritage research communities. In particular, there have been many fine examples of VR and AR incorporating 3D virtual reconstructions used for the communication of cultural heritage sites to non-specialists in the context of museums and visitors' centres.

There has been less research dedicated to exploring how 3D virtual reconstructions can be used in the research process itself, in particular to facilitate collaboration between specialists. A few examples of this are noted in the following section. In addition, within spatial disciplines such as archaeology, there is great potential for developing virtual reality tools that allow different specialists to construct simple 3D visualisations in a virtual environment themselves based on ongoing data collection.

In order to develop and test this tool, the site of Yeha, Ethiopia was chosen as it is an archaeological site currently under excavation within an Ethiopian-German cooperation project between the

---

<sup>1</sup> HafenCity Universität Hamburg, Labor für Photogrammetrie & Laserscanning, Henning-Voscherau-Platz 1, D-20457 Hamburg, E-Mail: alexanderpwalmsley@gmail.com, Thomas.Kersten@hcu-hamburg.de

<sup>2</sup> Deutsches Archäologisches Institut (DAI), Podbielskiallee 69-71, D-14195 Berlin, E-Mail: Mike.Schnelle@dainst.de

Authority for Research and Conservation of Cultural Heritage (ARCCCH), the Tigray Culture and Tourism Bureau (TCTB) and the German Archaeological Institute (DAI) and for which good remote sensing data already exists. In addition, a number of the research questions related to this project could benefit from a visual reconstruction of the prehistoric buildings at the site. As the excavations are still under way, this visualisation must however be able to evolve in line with the research.

## 2 Related Work

3D reconstruction, in particular coupled with Virtual Reality (VR), is being increasingly integrated into research and design processes in multiple industries. In the areas of architecture, urban planning and industrial design, commercial-grade tools such as Gravity Sketch have been available since 2014 (GRAVITY SKETCH 2022). In the areas of archaeological and heritage research, experimentation with 3D interactive tools goes back before the advent of commercially-available VR: REUTER et al. (2010) for example have described the development of a Tangible User Interface (TUI), called ArcheoTUI, that helps with the virtual reassembly of 3D-scanned archaeological artefacts. FORTE & WEBB CRAWFORD (2010) furthermore describes the development of a set of virtual, web-based collaborative 3D environments to be used for the purposes of archaeological research and education, based at the University of California Merced.

With the advent of commercially-available VR, some work has been done to integrate it into archaeological research (CASSIDY et al. 2019; TARI 2019). More recently, a joint project between BİMTAŞ, a company of the Greater Municipality of Istanbul, Turkey and the Photogrammetry & Laser Scanning Lab of the HafenCity University Hamburg, Germany, focusing on the 3D virtual reconstruction of the Selimiye Mosque in Edirne, Turkey, provided the opportunity to test VR as a platform for collaboration between specialists for architectural 3D modelling of monuments from the two countries (KERSTEN et al. 2017). A further collaboration between these two partners focused on the 3D laser scanning and virtual reality visualisation of the Rumeli Hisarı, an Ottoman fortress at the Bosphorus in Istanbul (TSCHIRSCHWITZ et al. 2019). In addition, VR has been explored as a tool for architectural historians and archaeologists for preserving a 3D digital copy of the Imperial Cathedral at Königsutter, Germany that can be explored with a headset and compared directly with the real-world environment through the integration of geo-located 360-degree panoramic photographs (WALMSLEY & KERSTEN 2020).

One of the most successful applications of VR in cultural heritage has been its use as an immersive and interactive tool for knowledge dissemination to non-specialist audiences, often in the contexts of museums. In this regard there are numerous successful use cases. BÜYÜKSALIH et al. (2020) describe for example the workflow for the production of a scanned 3D virtual environment of the İnceğiz caves, in the Çatalca district of Istanbul, Turkey, built in the Unity game engine and for use with the HTC Vive Pro headset. Further recent examples include the reconstruction of the town of Stade, Germany in the year 1620 for exhibition in the Schwedenspeicher Museum, Stade (WALMSLEY & KERSTEN 2019), or the reconstruction of Batavia (corresponding to the modern city of Jakarta), Indonesia in the year 1627 for the Westfries Museum, Netherlands (WESTFRIES MUSEUM 2022).

### 3 Yeha: the centre of the Ethio-Sabaeen community of Di'amat

An Ethiopian-German team has been undertaking collaborative research at the site of Yeha (Fig. 1) in order to investigate the phenomenon of the migration of southern Arabian populations to the Horn of Africa during the early first millennium BCE (GERLACH 2013; GERLACH 2014a; GERLACH 2017). It is around this time that culture-specific southern Arabian remains of Sabaeen influence begin to be found in south-eastern Eritrea and northern Ethiopia, indicating a cultural shift in this region (FATTOVICH 2009; GERLACH 2015; SCHNELLE 2014; SCHNELLE 2021). This influence is present from the ninth century BCE and is visible at numerous sites in the region in contexts relating to political, social, and religious life: the transfer of culture and techniques extended to the form of rule, the written and spoken languages, religion and cult practices, architecture, building techniques as well as artisanal crafts (GERLACH 2017). Through this process a sustained merging of Sabaeen and African elements was established: out of two diverse societies developed a single new culture that now bears the name “Ethio-Sabaeen” and that endured until the mid-first millennium BCE. The political and religious centre of the newly-established community called Di'amat is located at the contemporary site of Yeha in the northern Ethiopian province of Tigray. Around the middle of the first millennium BCE Di'amat ceases to exist. The best-preserved Ethio-Sabaeen building from Yeha is the Great Temple (mid-seventh century BCE) dedicated to the principal Sabaeen god Almaqah. Given the height of the exterior walls, which at some points are preserved up to the lower part of the roof, 14m above ground level, as well as the state of the rest of the surviving structure, the layout of the inner rooms can be reconstructed as a two-story hypostyle hall with cella walls made of wood and stone. The main entrance took the form of a six-column propylaea with a monumental doorway.

Only 200m from the Great Temple is situated the second monumental structure of the town, the Grat Be'al Gibri. With dimensions of nearly 60 x 50m this palatial complex of at least four stories was the largest wood-stone structure of eastern Africa and southern Arabia during antiquity. According to C14 dating the structure was first built around 800 BCE and thus is at the beginning of a long tradition of wood-stone architecture, which continues today with traditional Ethiopian architecture (SCHNELLE 2014).

Excavations in the Yeha necropolises have been looking at burial types and customs of the first millennium BCE. As of yet, all burials appear to be pit inhumations with two chambers and a tumulus-like superstructure. Due to plundering and the climate-induced poor state of conservation of the bones, it has been impossible to make further statements on the chronological arrangement of the finds (GERLACH 2017).

Excavations of the antique settlement should illuminate more of Yeha's urban structures as well as its different functional areas and provide material to establish a ceramic sequence. Only through settlement excavations will it be possible to grasp and chronologically situate the complex culture of Di'amat, with its indigenous and exogenous influences (GERLACH 2017).

Alongside scientific research activities, other activities currently taking place in Yeha are the restoration of the antique monumental structures, the development of the site for tourists, as well as the building of a local museum (GERLACH 2014b; TIGRAY CULTURE AND TOURISM AGENCY 2019; LEGESE & ALEMSHET 2019).



Fig. 1: Landscape around the settlement of Yeha, Ethiopia. Blue circle shows close-up of archaeological site

## 4 Methodology

The Great Temple had already been recorded in the autumn of 2009 by a joint team from the HafenCity University Hamburg and the German Archaeological Institute using terrestrial laser scanning and digital photogrammetry (LINDSTAEDT et al. 2011).

Digital Elevation Models (DEMs) at three separate resolutions were used to construct the topography of the site and its surroundings. Shuttle Radar Topography Mission (SRTM) data at a resolution of 30m (1/3 arc-second) was used to construct an area of 10.33km<sup>2</sup> surrounding the site. In order to use this data, it first had to be downloaded and processed in QGIS in order to extract a suitable region. 10.33km<sup>2</sup> was chosen as a suitable area through iterative testing as it seemed to be the smallest area of land in which the VR user could not see the horizon and therefore felt immersed in the landscape. It was also large enough to include several of the neighbouring mountains, which provide useful points for orientation within the virtual environment. The DEM was exported from QGIS as a 16-bit PNG at a resolution of 4096px × 4096px and imported into World Creator 2, a 3D landscape creation and texturing program developed by German development company BiteTheBytes GmbH. The principal advantage of using World Creator 2 as a middle step is that it allows for the procedural texturing of the landscape, which can then be imported directly into UE4. For the area immediately surrounding the site, DEMs of two different resolutions were used: a 10cm resolution DEM that included height information for trees, other vegetation, and buildings at the site; and a 50cm resolution DEM that had been cleaned of all this additional information and contained just information about the height of the land beneath (Fig. 2).

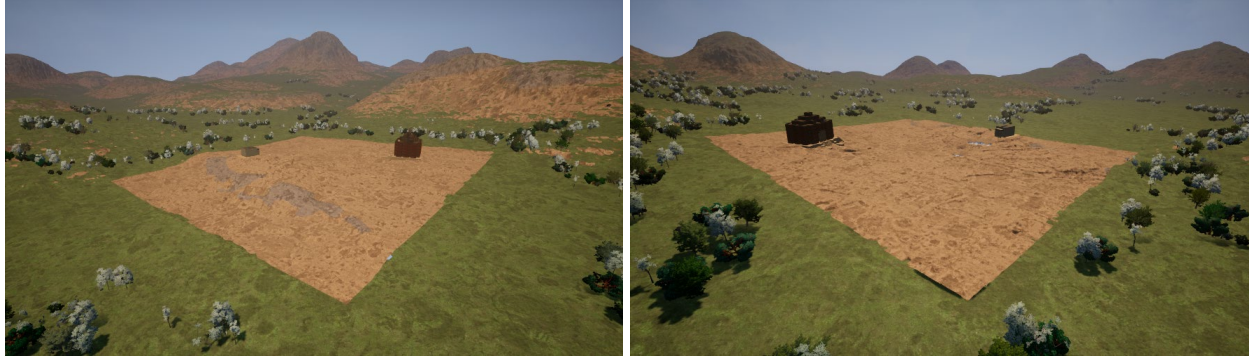


Fig. 2: Multiple-resolution system of landscapes. The background landscape (in green, with trees) is produced from the SRTM 30-meter resolution DEM, the foreground landscape (in brown) from the 50-cm resolution DEM generated by UAV photogrammetry

The area of both these regions was  $403.3\text{m}^2$ . Both DEMs were first loaded into QGIS for data extraction and re-export. The 10cm-resolution DEM was re-exported with dimensions of  $4033\text{px} \times 4033\text{px}$  (1 pixel every 10cm) and imported into UE4 as a landscape object. The origin of the virtual world was taken to be the origin of the UTM zone in which the DEMs were located (UTM 37N). The 10cm-resolution landscape object was then transformed in its x, y and z coordinates to its corresponding point within the UTM zone (this information was extracted from QGIS). In order to calculate the correct z-scale of the landscape, the following formula is provided in the Unreal Engine 4 documentation:

$$z^{\text{scale}} = 100(H^{\text{d}} - H^{\text{l}}) \times 1/512$$

In this formula,  $H^{\text{d}}$  refers to the altitude (in metres above sea level) of the highest point on the DEM,  $H^{\text{l}}$  refers to the altitude of the lowest point and  $z^{\text{scale}}$  refers to the scale factor within UE4. The figure of 1/512 is arrived at due to the fact that the UE4 reads 16-bit png values on a scale between -256 and +256 (512 values total). This allows us to arrive at a  $z^{\text{scale}}$  value for the 10cm-resolution DEM of 9.3163. The same workflow was also used for the 50cm-resolution DEM. This DEM was exported from QGIS with dimensions of  $806.6\text{px} \times 806.6\text{px}$  (1 pixel every 50cm). The  $z^{\text{scale}}$  was calculated to be 7.197.

In order to integrate the laser-scan data within the virtual environment, each model was first imported into a 3D program, Blender, and reduced to around 200 thousand polygons using the “Decimate” modifier. The model was then re-exported as an .obj file and imported into UE4. Since an .obj file retains the world position of the object in the vertex data, the origin can be set to the world origin within UE4 and the object then sits at the correct location in the world.

With the DEM-generated landscapes and laser-scanned models now sitting at the correct position within the virtual environment, the landscape object could then be adapted to reflect the historical height of the land, as revealed by the archaeological excavations. Using the native landscaping tools within UE4, the surface of the DEM-generated landscape was adjusted to the height of the laser-scanned models. The area of land in between the models was also adjusted so as to gradually slope from one model to another. In these areas then, the height of the land was extrapolated from the known data points of the laser-scanned models. Finally, the terrain overlaying the 3D model was made transparent in order to allow users to see the model properly (Fig. 3).



Fig. 3: Terrestrial scanning data (grey) of one of the archaeological trenches, integrated within the virtual environment in the lower height level of the past

Once the environment had been set up to accurately visualise the physical environment, a number of tools were programmed to allow users (primarily archaeologists and geophysicists) to create and visualise buildings and other historic and prehistoric structures in the environment according to archaeological data. In collaboration with project partners, a number of functionalities were prototyped and implemented. The following functions were programmed in UE4 using the native Blueprints C++ visual coding tools.

- i. Place new buildings on the terrain and define their dimensions: Pressing the menu button on the VR motion controller twice brings up the building placement menu. A user interface is also created to help the user pass in values for the width, length, height and rotation of the building using a slider. A ray-tracer allows the user to point at a point on the terrain and generate a building at that location when they press the trigger button on the VR motion controller. The numerical values for the different dimensions are also displayed alongside the sliders (Fig. 4)

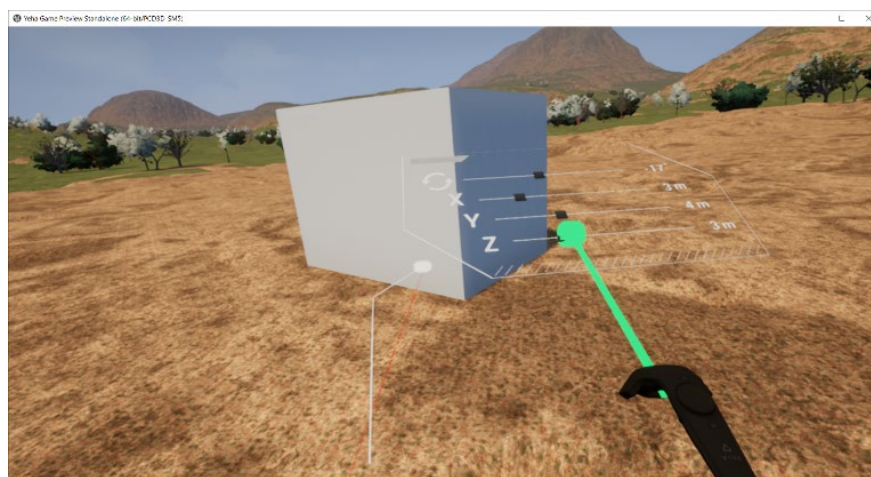


Fig. 4: User interface with sliders for the placement of the new buildings enabling the user to input values for the width, height, length and rotation of the building mesh



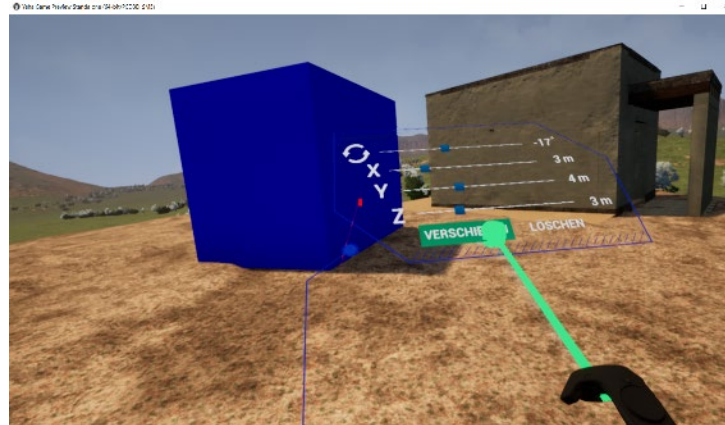


Fig. 5. User interface for editing the building, allowing the user to change the dimension parameters of the mesh, as well as move it to a new location on the terrain or delete it altogether

- ii. Allow the dimensions of these buildings to be edited after they have been placed: Once the building has been placed on the terrain, it can be further edited by pointing the VR motion controller at the mesh and clicking the trigger. This brings up an editing UI with the previously defined width, height, length and rotation displayed on sliders. These sliders can be further moved in order to change these parameters. While a building is being edited, it is highlighted dark blue to distinguish it from other meshes (Fig. 4).
- iii. Allow the building to be moved to a new location on the terrain or else deleted: At the bottom of the editing UI there are options that allow the user to delete the building (the building mesh is therefore removed permanently from the world) or move it to a new location. In order to perform this latter operation, the user must then point to a new location on the terrain and once again click the trigger button of the VR motion controller in order to place the building at its new location (Fig. 5).
- iv. Save the location of the buildings before closing the application and load it when starting again: If the environment contains buildings that the user would like to save before closing the app, the 'S' key on the keyboard allows them to save the environment in its current state. This state will be automatically loaded when the app is restarted. In order to delete the state, the 'D' key can be pressed during play, and the current state will be cleared from the cache (Fig. 6).



Fig. 6: Messages notifying the user that the data has been saved (left) and/or deleted (right)

- v. For the Almaqah Temple and Grat Be'al Gibri (Fig. 7), users should be able to view the different building phases: By pointing the VR motion controller at one of these buildings and clicking, the user is able to bring up the phase UI. This contains a menu allowing the user to replace the visible mesh with a mesh representing another phase in the building's life (Fig. 8). In this way the different phases can be directly compared within the same virtual environment.



Fig. 7: Reconstructed buildings in Yeha - Grat Be'al Gibri (left) and Almaqah Temple (right)

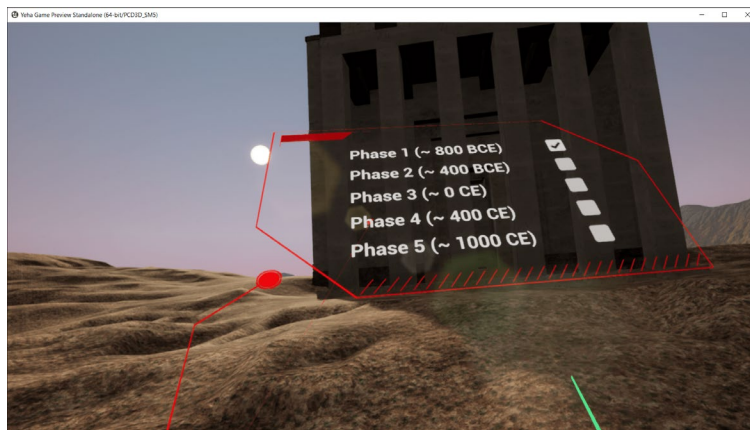


Fig. 8: Menu for selecting different building phases of the Almaqah temple by pointing the VR motion controller

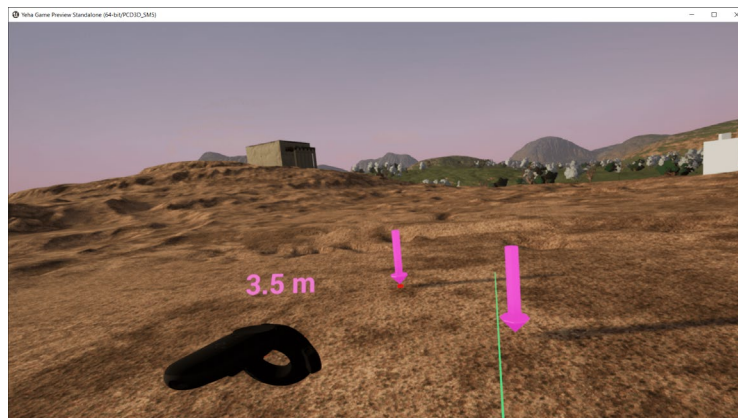


Fig. 9: Measurement tool, which consists of two anchor points placed on the terrain and the numerical value for the distance between them, displayed above the left controller

- vi. Measure the straight-line distance between two points on the terrain: Pressing the menu button on the VR controller once brings up the measurement interface (Fig. 9). Again, a ray tracer allows the user to target a location on the terrain to place the anchor point for the measurement. Clicking the trigger button once places this anchor point at the location and activate a second arrow to which the distance is calculated. The distance in metres appears as a numerical value above the left motion controller.

One function that had been envisaged but was not able to be implemented was one allowing the user to import new 3D models from a local or remote drive and integrate them into the virtual environment during play. This was deemed not to be possible due to restrictions on the way in which developers are able to use the Unreal Engine, which seems not to support real-time import of 3D models due to concerns surrounding software stability and security.



Fig. 10: The reconstructed archaeological site of Yeha with the Grat Be'al Gibri in the background and Almaqah Temple in the front indicating the present buildings in grey.

## 5 Conclusion and outlook

The VR reconstruction tool and associated functions for real-time 3D visualisation provide a starting point for the further development of VR tools for archaeological and architectural research. Further tests are already being conducted with data from another related site, that of Sirwah in Yemen, where there is a temple belonging to the same tradition as at Yeha. While there is a significant body of work dedicated to using VR in cultural heritage, much of this has focused on VR as a tool for knowledge dissemination, rather than as one for the research process itself. This paper addresses this gap and describes a number of prototyped tools that can be used by

archaeologists and other specialists themselves to visualise a site in real time. Further research might include tools to change the resolution of the landscape (for example by swapping out a low-resolution terrain for a high-resolution one), load in new 3D models in real time, and more detailed modelling tools for different kinds of structures. Finally, the archaeological site of Yeha has been reconstructed in virtual reality including the landscape of the environment, the two ancient buildings Grat Be'al Gibri and Almaqah Temple, and the present buildings indicated as boxes in grey (Fig. 10).

## 6 References

- BÜYÜKSALIH, G., KAN, T., ÖZKAN, G. E., MERIC, M., ISYIN, L. & KERSTEN, T., 2020: Preserving the Knowledge of the Past Through Virtual Visits: From 3D Laser Scanning to Virtual Reality Visualisation at the Istanbul Çatalca İnceğiz Caves. *Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, **88**, 133-146, <https://doi.org/10.1007/s41064-020-00091-3>.
- CASSIDY, B., SIM, G., ROBINSON, D.W. & GANDY, D., 2019: A Virtual Reality Platform for Analyzing Remote Archaeological Sites. *Interacting with Computers*, **31**(2), 167-176, <https://doi.org/10.1093/iwc/iwz011>.
- FATTOVICH, R., 2009: Reconsidering Yeha, c. 800–400 BC. *African Archaeological Review*, **26**(4), 275-290, <https://doi.org/10.1007/s10437-009-9063-3>.
- FORTE, B. & WEBB CRAWFORD, J., 2010: Participatory research in cyber-archaeology. CAA2009 - Making History Interactive. *Computer Applications and Quantitative Methods in Archaeology (CAA)*. Proceedings of the 37th International Conference, Williamsburg, Virginia, United States of America, March 22-26 (BAR International Series S2079). <http://dx.doi.org/10.15496/publikation-2888>.
- GERLACH, I., 2013: Cultural contacts between South Arabia and Tigray (Ethiopia) during the early 1st millennium BC: results of the Ethiopian-German cooperation project in Yeha. *Zeitschrift für Orient-Archäologie*, **6**, 254-277.
- GERLACH, I., 2014a: Migration Processes on the Northern Horn of Africa in the Early First Millennium BC. *Arabian and Islamic studies*, A. V. Sedov (ed.), A Collection of Papers in Honour of Mikhail Borishovich Piotrovskij on the Occasion of his 70th Birthday, Moscow, 100-120.
- GERLACH, I., 2014b: Yeha, Äthiopien: Site Management and Capacity Building. *e-Forschungsberichte des DAI*, **1**, 32-33.
- GERLACH, I., 2017: Neue Forschungen zur äthio-sabäischen Kultur, in: Steffen Wenig - Burkhard Vogt (Hrsg.), In kaiserlichem Auftrag: die Deutsche Aksum-Expedition 1906 unter Enno Littmann. *Ethnographische, kirchenhistorische und archäologisch-historische Untersuchungen, Forschungen zur Archäologie außereuropäischer Kulturen 3.3*, Wiesbaden, 355-390.
- GERLACH, I., 2015: Yeha, Äthiopien: Forschungen zu Migrationsprozessen am nördlichen Horn von Afrika. *e-Forschungsberichte des DAI*, **3**, 1-3.
- GRAVITY SKETCH, 2022: Think in 3D. Create in 3D. <https://www.gravitysketch.com>, last access 10.01.2023.

- KERSTEN, T., BÜYÜKSALIH, G., TSCHIRSCHWITZ, F., KAN, T., DEGGIM, S., KAYA, Y. & BASKARACA, A. P., 2017: The Selimiye Mosque of Edirne, Turkey - An Immersive and Interactive Virtual Reality Experience using HTC Vive. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, **42**(5/W1), 403-409, <https://doi.org/10.5194/isprs-archives-XLII-5-W1-403-2017>.
- LEGESE, T. & ALEMSHET, D., 2019: The Outstanding Heritages of Adwa, Ethiopia. *International Journal of Religious Tourism and Pilgrimage*, **7**(3):10, <https://doi.org/10.21427/ds65-9898>.
- LINDSTAEDT, M., MECHELKE, K., SCHNELLE, M. & KERSTEN, T., 2011: Virtual Reconstruction of the Almaqah Temple of Yeha in Ethiopia by Terrestrial Laser Scanning. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, **38**(5/W16), 199-203, <https://doi.org/10.5194/isprsarchives-XXXVIII-5-W16-199-2011>.
- REUTER, P., RIVIÈRE, G., COUTURE, N., MAHUT, S. & ESPINASSE, L., 2010: ArcheoTUI - Driving virtual reassemblies with tangible 3D interaction. *Journal on Computing and Cultural Heritage, Association for Computing Machinery*, **3**(2), 1-13, <https://doi.org/10.1145/1841317.1841319>.
- SCHNELLE, M., 2014: Monumentalbauten des 1. Jahrtausends v. Chr. in Yeha (Äthiopien) und Vergleichsbauten in Südarabien: Architektur als Spiegelbild von Kulturtransfer. *Zeitschrift für Orient-Archäologie*, **7**, 368-391.
- SCHNELLE, M., 2021: Observations on architectural features from the early 1st millennium BC in South Arabia and East Africa. *Contacts between South Arabia and the Horn of Africa, from the Bronze Age to Islam*, Ch. Darles, L. Khalidi, M. Arbach (eds.), in Honor of Rémy Audouin, *Sites et Cités d'Afrique Series*, Toulouse, 145-158.
- TARI, J.P., 2019: Deploying low cost virtual reality for archaeological research. Master's Thesis, University of British Columbia.
- TIGRAY CULTURE AND TOURISM AGENCY, 2019: Yeha – The Cultural Guide. <https://www.dainst.org/documents/10180/15360/Brosch%C3%BCre+Yeha+-+A+Cultural+Tourist+Guide/6f4bb7f9-dc41-4bd7-8892-7f562d73f2dc>, last access 10.01.2023.
- TSCHIRSCHWITZ, F., BÜYÜKSALIH, G., KERSTEN, T., KAN, T., ENC, G. & BASKARACA, A. P., 2019: Virtualising an Ottoman Fortress - Laser Scanning and 3D Modelling for the development of an Interactive, Immersive Virtual Reality Application. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, **42**(2/W9), 723-729, <https://doi.org/10.5194/isprs-archives-XLII-2-W9-723-2019>.
- WALMSLEY, A. & KERSTEN, T., 2020: The Imperial Cathedral in Königsutter (Germany) as an Immersive Experience in Virtual Reality with Integrated 360° Panoramic Photography. *MDPI Journal Applied Sciences, Special Issue Augmented Reality, Virtual Reality & Semantic 3D Reconstruction*, **10**, 1517; <https://doi.org/10.3390/app10041517>.
- WALMSLEY, A. & KERSTEN, T., 2019: Low-cost development of an interactive, immersive virtual reality experience of the historic city model Stade 1620. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, **42**, 405-411, <https://doi.org/10.5194/isprs-archives-XLII-2-W17-405-2019>.
- WESTFRIES MUSEUM, 2022: Batavia 1627 in virtual reality. Hoorn, Netherlands. <https://wfm.nl/batavia-1627vr>, last access 10.01.2023.

# An Approach to Develop a Forestry Data Space Using Geoinformation and Remote Sensing Data for Forest Management and Protection

NINA KRUEGER<sup>1</sup> & STEPHAN PLABST<sup>1</sup>

*Abstract: The digital transformation strongly affects the field of forest management and protection and provides a lot of potential for improvement in many decision-making processes. However, in the actual forestry ecosystem, despite of numerous solutions, often most of them are not compatible. In this study, we outline a preliminary forestry data space concept, which resulted mainly from the FutureForest project “KI-Einsatz bei Waldzustandsanalyse, der Bewertung zukünftiger Waldentwicklung sowie Entscheidungsvorbereitung zum klimaangepassten Waldumbau” (FF.ai) and describe the technical components. The implemented prototype would try to address the main existing obstacles such as data restriction, complex access and data gaps to extend acceptance of digital technologies in forestry and remote sensing community. Furthermore, a forestry data space as part of an extended digital ecosystem with example scenarios will help to advance digitalization in forestry.*

## 1 Introduction

In our surrounding environment, transformation in digital form is gaining speed and data plays an important role in this transformation, in terms of functionality and sharing. Moreover, the amount of acquired data with high value is rapidly growing. However, make an optimal use of this huge available data is still lacking due to the restrictions and complicated access rules, limited or nonexistence trust between data providers and data users, and data gaps (CURRY 2016). In this regard, to address the mentioned issues, new developments in the area of data infrastructures and their future is happening under the name so-called data spaces in Germany and worldwide. Initiatives such as International Data Space (IDS) and GAIA-X community are leading the design principles of a data space model under important criterions such as data sovereignty, governance, interoperability, portability and trust between participants. Data Space (DS) stands as a decentralised infrastructure for trustworthy, qualified, transparent data sharing and exchange based on commonly agreed principles and rules. It enables many parties to access, use and share their data with each other efficiently and securely. In contrast, to the current data infrastructures with mostly uncontrolled access/use of data, a DS attempts to guarantee confidentiality and security (FRANKLIN et al. 2005). In case of removing restriction and making data open, however, the owners of the data are often exposed to the danger of losing control of their data and thus the strategic value of their data resources without gaining anything as exchange. In consequence, the sharing of data is still only taking place in a limited extent and the potential of DSs remains not explored.

---

<sup>1</sup> M.O.S.S. Computer Grafik Systeme GmbH, Hohenbrunner Weg 13, D-82024 Taufkirchen,  
E-Mail: nkrueger@moss.de

Nowadays, DSs developments are focusing on the implementation of the European data strategy in different sectors. While the first use cases are already available in industry, a deeper understanding about the design and properties of a sector specific DS such as forestry with its own specific remote sensing data requirements and regulations is necessary. In this regard, a DS for forest management purposes is aiming at creating and maintaining a single environment for forest relevant data and processes to shared and exchanged across different participants efficiently and securely. This is a challenging goal due to missing reference data and lacking or varying quality of the input data. Therefore, it is necessary to know and understand the concept and architecture model of a DS to adapt the requirements of it to the key aspects of the users and applications in geoinformation and remote sensing community for forest management as well.

The FutureForest<sup>2</sup> (FF.ai) as an ongoing project (until 2025), aims to develop a DS concept in the field of forestry using remote sensing data and Artificial Intelligence (AI) techniques in order to support the climate-adapted decision making for forest management and planning in Germany. The project consortium members are FU Berlin (FUB), Chair of Ecosystem Dynamics and Forest Management at the Technical University of Munich (TUM), M.O.S.S. Computer Grafik Systeme GmbH (MOSS) and wetransform GmbH (WE). Furthermore, the conceptual prototype DS based on forestry-relevant data explores harmonization and simple access regulations including a certification and authorisation stage for data and participants. The available technologies and scientific approaches in the community used as base line and extended to the forestry relevant user scenarios. Based on the conducted experiments during the project, the developed and later implemented prototype will create a road map aiding future forest management by enabling data providers to sharing data while preserving sovereignty on the secure exchange of data in the private and public sector.

## 2 Problem Statement and Motivation

Forests are changing continuously and playing an essential role in our environment. In the recent years, however, different natural hazards such as storms, fires and bark beetles had great impacts on the actual and future stands of forests. The decision of natural resources managers regarding the future of forest directly connected to the geospatial and remote sensing data in aids of monitoring forests in large scales. Therefore, an accurate, transparent, secure, and shared data ecosystem of forests including biophysical parameters are required to support sustainable forest management. Next to the rapidly rising remote sensing data, other issues regarding a secure shared data environment, which can be accessible and profitable for data provider or data user, are gaining importance. Additionally, it is clear that the remotely sensed data is a dynamic matter in order to support the long-term forest monitoring purposes. Therefore, a DS for forestry management purposes that has access to geoinformation and remote sensing data from various open or private resources, regularly updated, secured, based on the simple rules, providing products and services to the participants is necessary.

---

<sup>2</sup> <https://future-forest.eu/>

### 3 Concept

After a short introduction to DS initials and the goal of FF.ai project, this section describes the main components of a DS known as the soft infrastructure (3.1), connectors and their role as a gateway in this project (3.2) and digital identity (3.3).

#### 3.1 Components, Information Model and Governance

In this section, the structure and main components of a DS individually explained (Fig. 1). Additionally, a detailed description provided to understand the phenomenon of data sovereignty and the architecture as a federated approach for the sovereign exchange of data (OTTO 2022; BADER et al. 2020; JARKE et al. 2014). In following, the connectors and their roles in DS according to the current IDS reference architecture and finally an example scenario, which specially defined for forestry application, described.

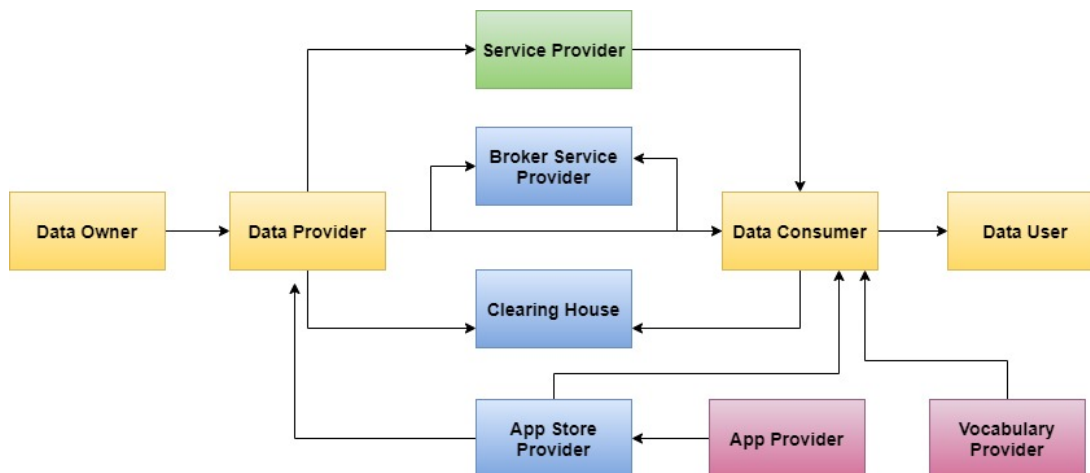


Fig. 1: The soft infrastructure of a DS and the main components

##### 3.1.1 Data Sovereignty

The key factor data sovereignty stands as a precondition for trustworthy data exchange and supports aspects such as identity, trustworthiness of participants. Moreover, it controls data access and usage as well.

##### 3.1.2 Data Governance

In DS fundament, governance plays an important role as a rule-based mechanism (terms and conditions), which is responsible to put data visibility, transparency, and sovereignty in order. As an example, a data can be open for all participants or limited to a specific group based on the allowance of provider.

##### 3.1.3 Data Owner/Provider

The data owner stands as a party who defines and/or exercises control over the data by defining policies and allowing access to data. It should have the technical capability and the responsibility in case of usage contracts and policies. In the other hand, the data provider is responsible for providing the data for exchange between an owner and a consumer including software components



that are compliant with the IDS Reference Architecture Model for this purpose. The data provider and data owner not necessarily, but in most scenarios are identical.

#### 3.1.4 Data Consumer/User

It receives data from a data provider and before initializing a request for data, searches for existing datasets by making an inquiry at a Broker Service Provider (see 3.1.5). The data user is the entity that has the legal right to use the data of a data owner as specified by the usage policy. In most cases, the data user is identical with the data consumer.

#### 3.1.5 Broker Service Provider

This component acts as a mediator between data providers and users requesting for data. A Broker provides functions, for data owners to publish their data sources, data users to search through the data sources and to make agreements on the control and use of specific data.

#### 3.1.6 Clearing House

This component keeps/logs all activities performed in the course of a data exchange. After a data exchange completed, or parts of it, both the data provider and the consumer confirm the data transfer by logging the details of the transaction at the Clearing House. Next to keeping record of exchange transactions, the clearing house reports on the search for data sources and on exchange transactions as well.

#### 3.1.7 Service Provider/APP Provider

The IDS promotes the development of a business ecosystem in which participants may develop software (especially services) and make this software available via the App Store. Apps development direction should be compliant with the system architecture of the IDS to be used in the IDS relevant use cases including certifications (MENZ et al. 2019) in order to increase trust in these applications. In this project, a web GIS service further extends based on the WEGA<sup>©</sup> software technology from MOSS (M.O.S.S. 2019) and a decision support app develops by WE to support the forest management tasks.

#### 3.1.8 Vocabulary Provider

In order to better define how one party owns data, specific vocabularies can be defined and made available for all the DS participants. This component of DS manages and offers vocabularies (i.e., ontologies, reference data models, or metadata elements) in order to annotate and describe datasets.

### 3.2 Data Space Connectors

The Connector ensures that participants maintain sovereignty over the data and at the same time, functions as an interface between the internal systems of the DS participants and the DS itself. Depending on the configuration, the Connector hosts a variety of system services ensuring, for example, a secure communication, enforcement of usage policies upon exchanged content, monitoring, and logging of content transactions for clearing purposes. Here, a detailed evaluation of available connectors and their performance is out of scope.

### 3.3 Digital Identity

Certification is an important element in the DS to establish a trustworthy communication. The stage of certification, authorisation based on user and role type in DS defined as digital identity to provide controlled and restricted access via rules and pre-defined regulations. The Identity Provider shall offer a service to create, maintain, manage, monitor, and validate identity information of and for participants in the DS. This is to avoid unauthorized access to data. Further investigations on security and identity management is out of focus of this paper.

## 4 Application Scenario

In forestry, there are variety of potentials for improvement in case of planning and decision making with respect to the climate adaptation. To benefit from an optimal operational management system supported by software services, all data acquired for decision-making processes must be existed in digital form (LENZERINI 2002). However, for digitalization, the complex models, with lots of parameters and a high amount of required data is challenging. Moreover, there are many different, partly isolated services and systems with redundant or missing data, processes, and software solutions, which cannot be connected with each other due to the lack of interoperability of data, different interfaces and regulations. Furthermore, specific challenges of the forestry such as offline capabilities, use of data locally, or lack of minimum IT infrastructure in private sector, have to be considered for any further integration.

In the first step, a conceptual forestry DS should provide an infrastructure for efficiently supporting forestry interactions and work processes with services and information, provide boundaries and specifications for flexible adaptation of needs from different participants (UL HASSAN et al. 2020). The FF.ai project team combines the expertise of science and research with the interests of the private sector to develop a concept for such a DS while placing challenges like data availability, sovereignty and interoperability in the focus. The ongoing development on a DS prototype in FF.ai project using IDS reference architecture model and connectors at the current state for use cases presented in Fig.2. Each digital DS initially formed by two or more participants. In this project, these participants are foresters, GIS consultants, and research organizations. Three actors involved in a scenario that focuses on use of AI techniques to provide information that can support climate-adapted forest management are:

- The foresters as the data owners,
- The research organisations as the data users and at the same time providers,
- Third-party system or service providers (GIS services, DSS App etc.).

In the proposed scenario, three different participant groups defined to involve and have access via connectors to the DS, which either provide or collect data, use or evaluate or process data for AI based analysis and provide services/apps. Here, the DS approach can offer supporting services via connectors and at the same time protect the forester's data, as not all data necessarily needs to be free for everyone. Furthermore, data sovereignty plays an important role while besides open data; most of the data in forestry acquired from different governmental sources and private owners. In order to optimize the access cycle, the relevant data must be complete and shared in agreed quality by participants. Thus, the participant as forester can get information about his/her current forest

status and thus identify possible problem areas and ask for consultant. Alternatively, a GIS consultant who wants to evaluate a catastrophic situation in forest and provide the owner solutions can use the advantage of connecting to the DS. The connector as a gate provides a secure and authorised access to the DS in order to meet the requirements such as collect the necessary data to do the GIS analysis and the use the results of a trained AI model. As another application example for the DS and corresponding services, is the evaluation of ecological and economic sustainability via the climate change adaption. In forestry, a balanced and appropriate climate adapted cycle forms the core of the efficient, productive, and sustainable planning. In our project, a participant from scientific organisation as an AI expert uses the DS potential to access remote sensing data such as Sentinel 2, aerial imagery etc. for training and optimization of an AI process in order to provide tree species map. The aim, which realistically achievable is by linking different forest areas, remote sensing data and sensor systems records together.

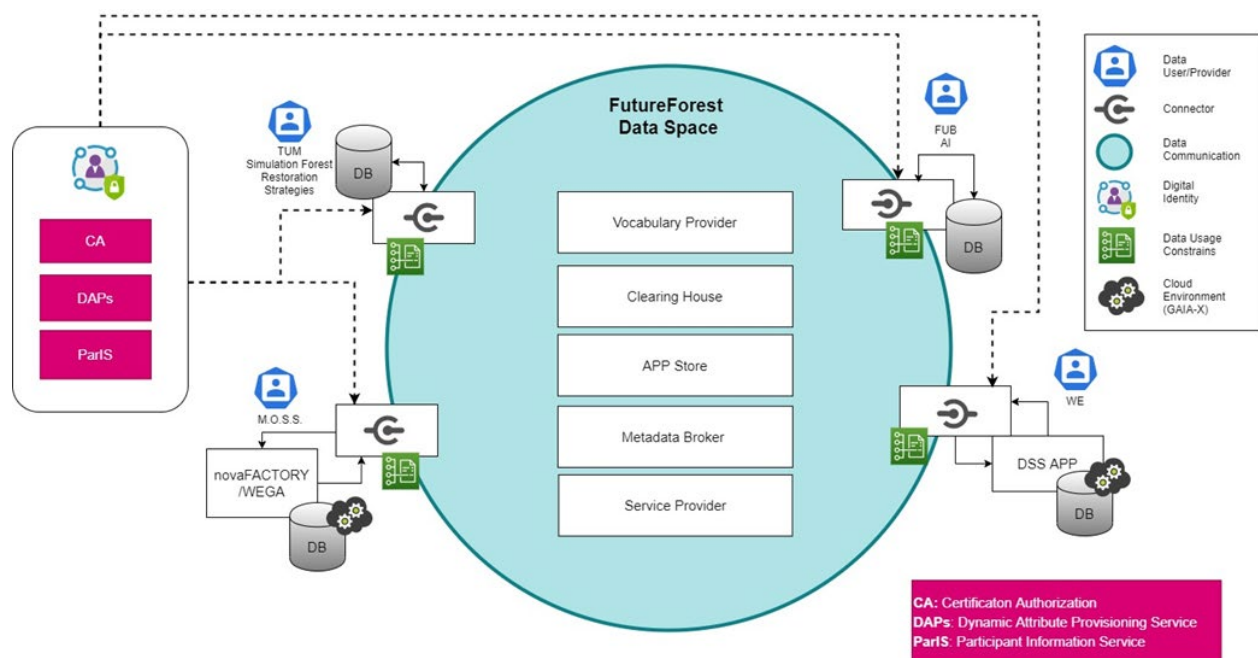


Fig. 2: The developed DS concept in FF.ai project using IDS reference architecture model (for detailed explanation of components see section 3).

Finally, the DS services offer the participants for decision-making aids through an integrated web GIS platform based on the software technology of *novaFACTORY*<sup>®</sup> and *WEGA*<sup>®</sup> by MOSS (M.O.S.S. 2019) and a decision support app in development phase by WE.

## 5 Summary and Outlook

In this study, we have motivated and presented the concept of a prototypic forestry DS, which can greatly advance digitalization in forestry. To do so, the proposed concept, takes the concepts of the IDS standardisations, adapts, and extends them with solutions for the forestry applications. However, there are still uncertainties especially in case of certification and authorisations possibilities with respect to the remote sensing data, which requires further investigations. Therefore, the association and participation from remote sensing and forestry community during this project is very important in order to develop the appropriate concept.

## 6 Acknowledgement

The three-year research project is funded by the German Federal Ministry for the Environment, Nature Conservation, Nuclear Safety and Consumer Protection (BMUV) under the funding program " KI-Leuchttürme für Umwelt, Klima, Natur und Ressourcen " (funding code: 67KI21002A). The BMUV is supported by Zukunft - Umwelt - Gesellschaft (ZUG) GmbH as project executing organization.

## 7 References

- BADER, S. R., PULLMANN, J., MADER, C., TRAMP, S., QUIX, C., MÜLLER, A. W., AKYÜREK, H., BÖCKMANN, M., IMBUSCH, B. T., LIPP, J., GEISLER, S. & LANGE, C., 2020: The international data spaces information model - an ontology for sovereign exchange of digital content. Int. Proceedings of International Semantic Web Conference (ISWC), Athens, Greece, 176-192, [https://doi.org/10.1007/978-3-030-62466-8\\_12](https://doi.org/10.1007/978-3-030-62466-8_12).
- CURRY, E., 2016: The big data value chain: Definitions, concepts, and theoretical approaches. In J. M. Cavanillas, E. Curry, & W. Wahlster (Eds.), New horizons for a data-driven economy, Springer International Publishing, 29-37, [https://doi.org/10.1007/978-3-319-21569-3\\_3](https://doi.org/10.1007/978-3-319-21569-3_3).
- EUROPEAN COMMISSION, 2021: Results of the new European Data Market study 2021-2023 online at <https://www.digital-strategy.ec.europa.eu/en/library/results-new-european-data-market-study-2021-2023>, last access 11.09.2022.
- FRANKLIN, M., HALEVY, A. & MAIER, D., 2005: From databases to dataspace. ACM SIGMOD Record, **34**(4), 27-33, <https://doi.org/10.1145/1107499.1107502>.
- GAIA-X, 2021: IDSA Reference Architecture Model, Version 4.0. <https://internationaldataspace.org/ids-ram-4-0/>, last access 10.12.2022, <https://www.bmwi.de/Redaktion/EN/Dossier/gaia-x.html>, last access 20.08.2022.
- JARKE, M., JEUSFELD, M. & QUIX, C., 2014: Data-centric intelligent information integration -from concepts to automation. Journal of Intelligent Information Systems, **43**(3), 437-462.
- LENZERINI, M., 2002: Data integration: A theoretical perspective. In Proceedings ACM-PODS, 233-246.
- MENZ, N., RESETKO, A. & WINKEL, J., 2019: IDS certification explained. <https://internationaldataspace.org/download/16456/>, last access 13.01.2023.

- M.O.S.S., 2019: Umsetzung eines Digitalen Lagers für Geodaten mit novaFACTORY. M.O.S.S. Computer Grafik Systeme GmbH, <https://t1p.de/nFdigitaleslager20>, last access 20.01.2023.
- NAGEL, L., 2021: Design principles for data spaces. <https://design-principles-for-data-spaces.org/>, last access 11.09.2022.
- OTTO, B., 2022: A Federated Infrastructure for European Data Spaces. Communications of the ACM, **65**(4), 44-45, <https://doi.org/10.1080/10580530.2015.1044344>.
- UL HASSAN, U., OJO, A. & CURRY, E., 2020: Catalog and entity management service for internet of things-based smart environments. In Real-time Linked Dataspaces, Springer, 89-103, [https://doi.org/10.1007/978-3-030-29665-0\\_6](https://doi.org/10.1007/978-3-030-29665-0_6).

# Transferability of Deep Learning Models for Land Use/Land Cover Classification

MAXIMILIAN HELL<sup>1</sup>, MELANIE BRANDMEIER<sup>1</sup> & ANDREAS NÜCHTER<sup>2</sup>

*Abstract: Deep learning models in remote sensing are often trained once for benchmarking their results and not further applied to new domains or newer data. In this study, we test five previously developed DeepForest model variations on new data for land use and land cover classification. The models were pre-trained for this task on a multi-modal and –temporal data set from 2018 covering parts of the Amazon rainforest. The data comprises a twelve-month time-series of Sentinel-1 SAR data combined with a single Sentinel-2 multispectral scene. Classification maps from the MapBiomas Brazil project are used as label data to assess classification performance. The DeepForest classification models are able to classify the test scene from 2020 with up to 83.84% overall accuracy, producing reliable land use and land cover maps.*

## 1 Introduction

New deep learning models for land use and land cover classification are published frequently, mostly without any further studies of their potential transferability to new satellite data or other areas. In a previous study, we trained five variations of our newly proposed DeepForest model (CHERIF et al. 2022) for land use and land cover classification in the Amazonas region of Brazil. The models leverage on the multi-modal and –temporal aspects of the input satellite imagery, consisting of a multi-spectral Sentinel-2 image and a twelve-month synthetic aperture radar (SAR) Sentinel-1 time-series. One major challenge for continuous landcover mapping is the transfer of the process to new data. The goal of developing the DeepForest models was to provide architectures that are capable of generalizing well, not just on validation data but also on new, large datasets. Thus, these models will be tested in the same study area, but on more recent satellite data and their respective labels. Testing the transfer of these models to the new data will show if they learned a generalized representation and are able to be further employed in a land use and land cover classification workflow or whether larger amounts of data and/or better labels are necessary to achieve the ultimate goal of an automatized workflow.

## 2 Materials

### 2.1 Study Area

The study area is located in the region Amazônia Legal in Brazil. This socio-geographic region comprises nine of Brazil's 26 states and contains the Amazonas basin and rainforest. Testing the performance of the trained deep learning models was performed on a mosaic of satellite data

---

<sup>1</sup> Technische Hochschule Würzburg-Schweinfurt (THWS),  
E-Mail: [maximilian.hell, melanie.brandmeier]@thws.de

<sup>2</sup> Julius-Maximilians-Universität Würzburg, E-Mail: andreas.nuechter@uni-wuerzburg.de

covering an area of 35,800 km<sup>2</sup> in the state of Mato Grosso, shown in Fig. 1. The satellite data was processed as described in Sec. 2.2.

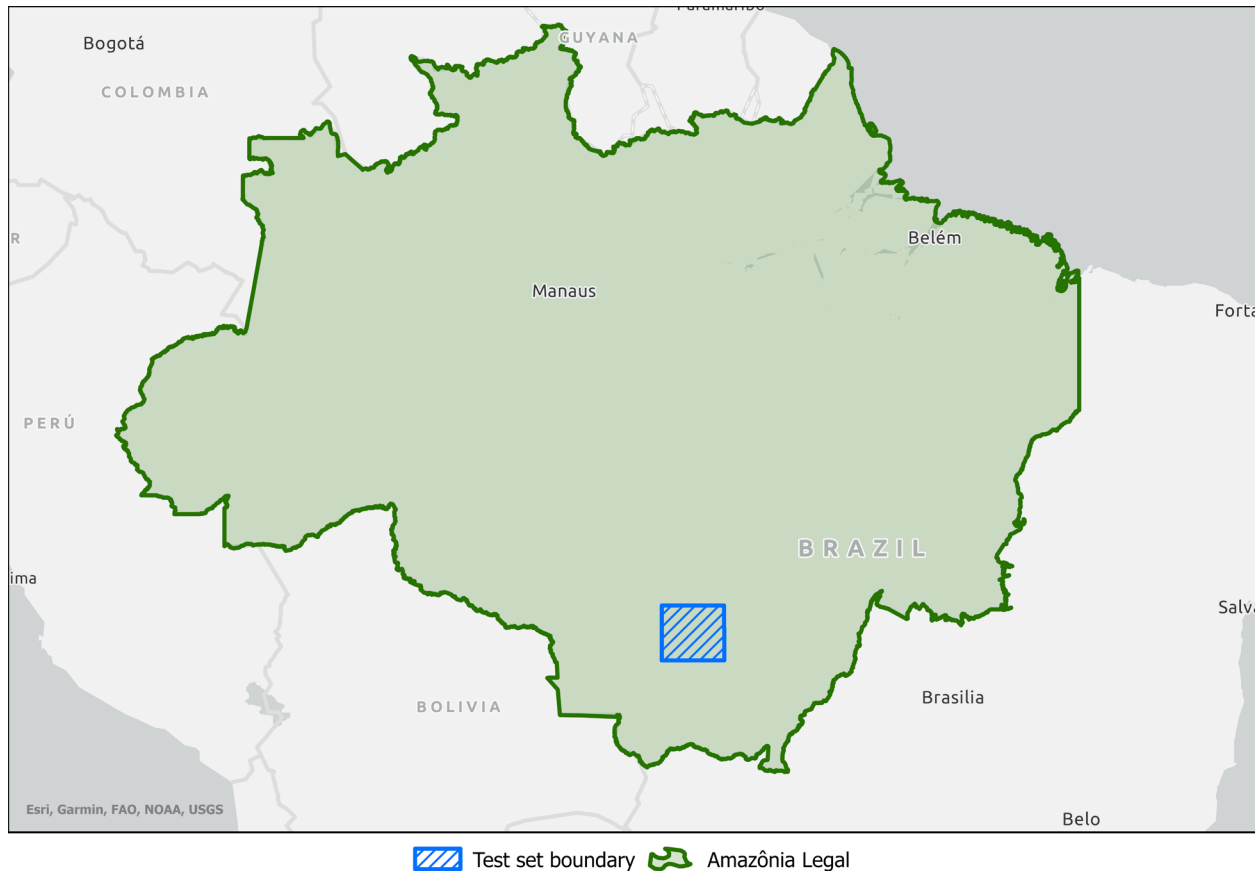


Fig. 1: Study area in the Amazon basin in Brazil. The footprint of the used satellite data mosaic is marked in blue.

## 2.2 Data

The satellite data used for the land use and land cover classification was preprocessed in the same way as described in the prior study (CHERIF et al. 2022). The data is comprised of a stack of a 12-month Sentinel-1 time-series and a single image of a Sentinel-2 scene. All data was captured in the year 2020, as this was the year with the most current label data available.

The Sentinel-1 SAR time-series contains twelve captures, one from each month, in VV+VH dual polarization, resulting in 24 bands. All scenes were captured in interferometric wide swath (IW) mode and acquired as Level-1 GRD products. The imagery was then further preprocessed using ESA's SNAP software (ESA 2023). The correction workflow (FILIPPONI 2019) comprises the following steps:

1. Precise orbit correction
2. Thermal noise removal
3. Calibration
4. Speckle Filtering (using the Refined Lee filter)
5. Conversion to the dB scale

The Sentinel-2 scene was chosen to have minimal cloud coverage and with an acquisition time in the middle of the year (June—August). In total, four image tiles were used, captured in the same pass of the satellite and spatially intersecting the Sentinel-1 time-series. The imagery was acquired from the Copernicus Hub as Level-2A product (Bottom-of-Atmosphere reflectance) with most of the atmospheric influences corrected. This product consists of 13 bands, of which only 10 are used. The three omitted bands have a pixel spacing of 60m×60m and are primarily utilized in atmospheric applications. The remaining bands with a pixel spacing of 10m×10m and 20m×20m were resampled to match the highest resolution of 10m×10m using the nearest neighbor method. All 34 bands, composed of Sentinel-1 bands and 12 Sentinel-2 bands, were then stacked and mosaicked to represent one image in a geodatabase using ArcGIS Pro (ESRI 2023).

As ground-truth labels, we used data from the MapBiomass project (MAPBIOMAS 2022). The project generates land use and land cover classification maps based on Landsat data. The published maps have a spatial resolution of 30m×30m, like the underlying satellite data. They publish iterations of their classification techniques as collections. We used Collection 6, as this was the most current version at the beginning of the project. The most recent year for a classification map within this collection is 2020. This collection uses a classification scheme, which comprises 26 different land use and land cover classes. However, only 20 of these classes are present in the Amazonas region.

The DeepForest models used for the transfer study were trained with label data from MapBiomass Collection 4. Data from this collection is provided up to the year 2018. Furthermore, the classification scheme changed from Collection 4 to Collection 6 with no apparent backwards compatibility, especially in the agricultural usage classes. The label map had to be reclassified to match the labels output by the models. To extract a class remapping scheme, the Collection 6 classification map from 2018 was compared to Collection 4 from the same year, further described in Section 3.2.

### 3 Methodology

#### 3.1 Deep learning models

The models used in this transfer study are pre-trained variations of the DeepForest models (CHERIF et al. 2022). These models were trained on a combination of a Sentinel-1 time-series and a single Sentinel-2 capture, as described in Section 2.2. The training satellite data was captured in the year 2018 and the labels from the MapBiomass project are within Collection 4 of the same year.

The models are grouped according to their respective data fusion approach: The DeepForest-1 family of models uses an early fusion approach and contains three variations: DF1a, DF1b, and DF1c. In these models, both input data modalities are processed together and in parallel. The DeepForest-2 models, DF2a and DF2b, use two different data streams in the network to process the SAR time-series and optical data separately. The learned representations are then fused for the final classification step. All models use convolutional long-short-term memory [ConvLSTM] (SHI et al. 2015) as a defining building block to efficiently learn the spatio-temporal relationships of the SAR time-series data.



### 3.2 Transfer of the classification scheme

The MapBiomass Collection 4 used to train the original models provided classification maps up to the year 2018. Based on this scheme, the pre-trained models are able to differentiate between thirteen land use and land cover classes: *Forest Formation*; *Savanna Formation*; *Forest Plantation*; *Wetland*; *Grassland*; *Other non Forest Formation*; *Pasture*; *Annual and Perennial Crop*; *Semi-perennial Crop*; *Urban Infrastructure*; *Other Non-Vegetated Areas*; *River, Lake and Ocean*; and *Mining*.

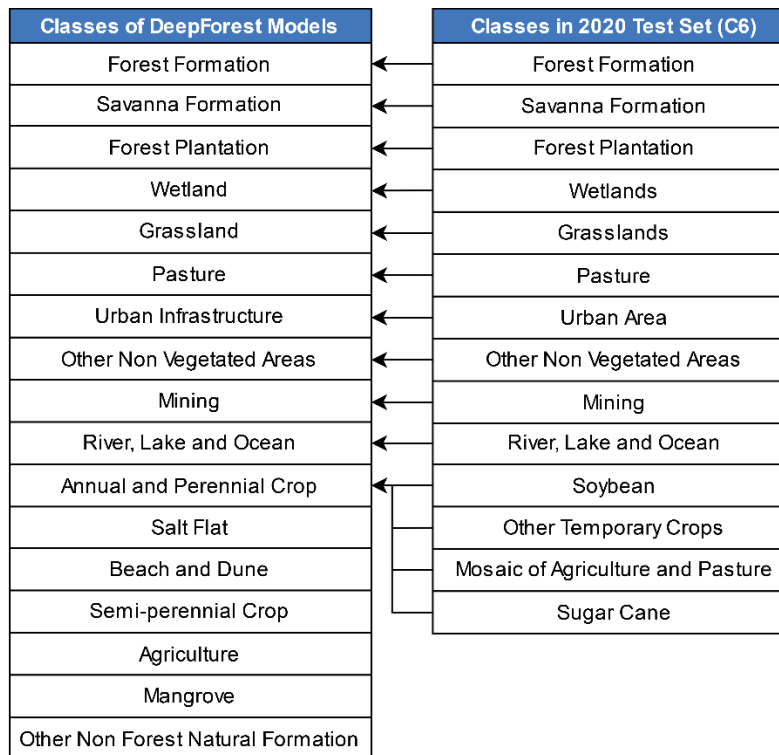


Fig. 2: Remapping from the classification scheme of Collection 6 to Collection 4 of the land use and land cover classes present in the new test set.

However, the data used for our transfer study was captured throughout 2020. The latest MapBiomass iteration at the start of the data processing was Collection 6 with an altered classification scheme. Thus, some re-mapping of the classes to the older scheme had to be conducted prior to inference and accuracy assessment. New (sub-)classes were introduced to the more recent collection and some classes were completely changed. Most significant was the change of the agricultural classes. Collection 4 used the distinction between *Annual and Perennial Crop* and *Semi-perennial Crop*. In contrast, Collection 6 distinguishes between *Perennial Crop* and *Temporary Crop*, with three and four further subclasses, respectively. These two macro and seven micro classes had to be reclassified into the two agricultural classes of Collection 4. The label maps covering the test area from 2018 of both collections were compared, to achieve this mapping from one scheme to the other. The *Soybean* class, contained in the *Temporary Crop* macro class of Collection 6, matched almost fully (98.76% of all pixels) with the *Annual and Perennial Crop* class. The class *Other temporary Crops* was not as unambiguously (67.38%)

ascribed to this class. *Sugar Cane* (also within in the *Temporary Crop* class) and *Mosaic of Agriculture and Pasture* were also mapped to this class. It is also noticeable that the class *Wetland* of Collection 6 is not present in this area in Collection 4, although this class exists in that scheme. The majority of the *Wetland* pixels (53.17%) map back to the *Savannah Formation* class. Figure 2 shows the remapping of the classes present in the test set to the classification scheme of Collection 4, on which the DeepForest models are classifying.

### 3.3 Inference on the test data

All five pre-trained model variations were used to classify the test scene (Fig. 1). The resulting maps were then compared pixel-wise to the label data from MapBiomass Collection 6 remapped to Collection 4, as described in Sec. 3.2. These label maps were registered to the satellite data and upsampled from 30m×30m using nearest neighbor sampling to match the pixel spacing of 10m×10m of the test data. All analysis was performed within ArcGIS Pro to make the best use of the mosaic dataset.

## 4 Results

In the original study (CHERIF et al. 2022) all models reached overall accuracies (OA) of at most 74.4% (DF1c) on the test set. On the new data set, all models performed better with respect to the OA. DF1c reached the highest score, with 83.84% of all pixels classified correctly. All models perform at least 8.9 percentage points higher than on the originally trained data, as shown in Tab. 1.

Tab. 1: Overview of the overall accuracies on the test set of the previous study and the new data set

Model	OA in CHERIF et al. 2022 [%]	OA on the new test set 2020 [%]
<i>DF1a</i>	74.3	81.57
<i>DF1b</i>	72.9	82.20
<i>DF1c</i>	74.4	83.84
<i>DF2a</i>	70.9	79.85
<i>DF2b</i>	69.0	80.84

However, three of the thirteen classes the models were trained on are missing in the new dataset. These classes are *Mining*, *Other non Forest Formation*, and *Semi-perennial Crops*. The classes *Forest Formation* and *Annual and Perennial Crop* are the majority land cover classes, with 32.4% and 37.2% of the test data set, respectively.

A qualitative assessment of the results by visual inspection should be conducted, when producing classification maps. Fig. 3 shows the ground truth labels of the MapBiomass project compared to the resulting map of the DF1c model, which achieves the highest overall accuracy in the quantitative assessment. The classification of *Annual and Perennial Crop* and *Pasture* is very similar to the ground truth. Some of the small *Savanna Formation* patches are missing in the classification. The DF1c model classifies much less area as *Urban Infrastructure* (dark red) than given as ground

truth. When comparing this to the satellite image in Fig. 4 the model classifies the underlying land use and land cover classes well.

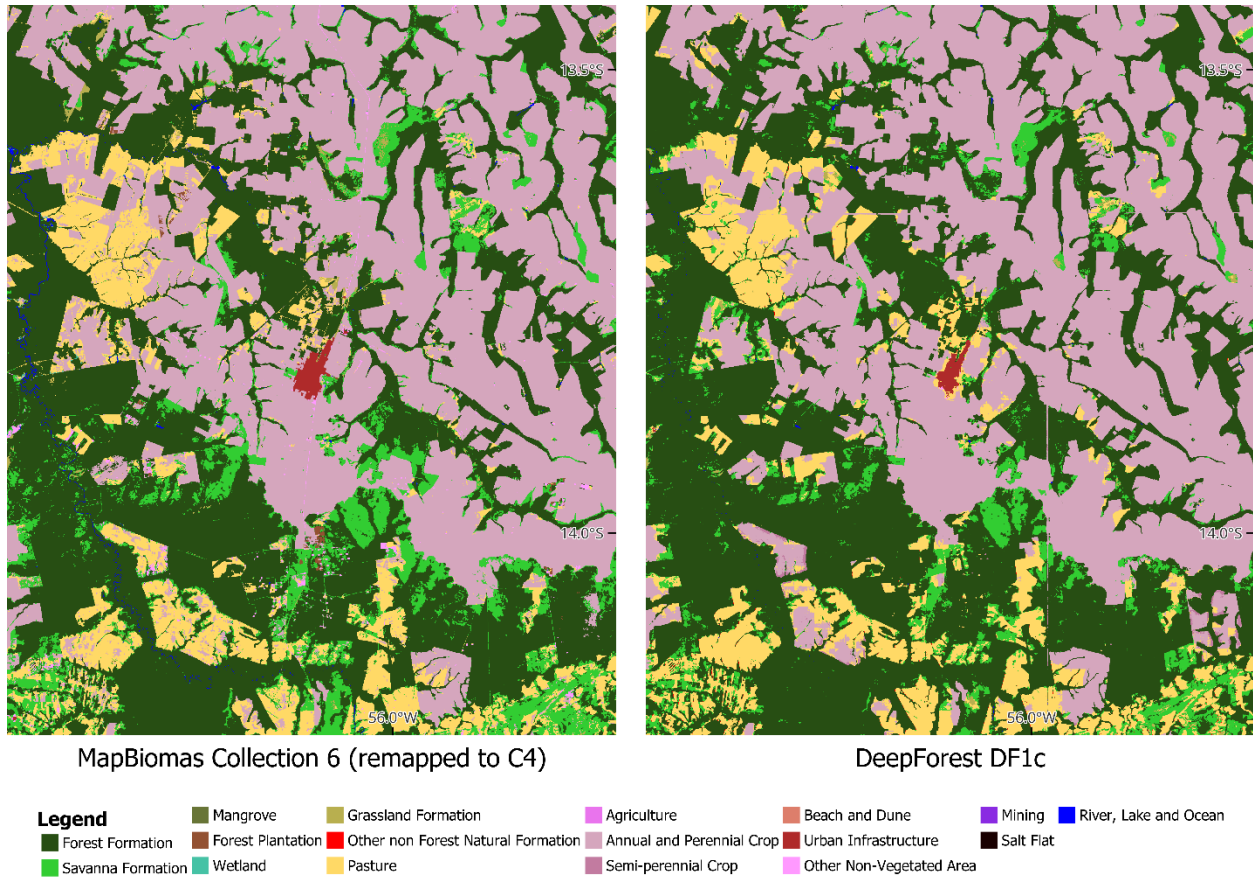


Fig. 3: The resulting land use and land cover map produced by the DF1c model (right) compared to the ground truth of MapBiomass (left).

Figure 4 shows a satellite base image, as well as one of the models from the DeepForest-2 late fusion family (DF2a) with the same spatial extent. The classified map shows a huge overlap with the ground truth in the agricultural and *Forest Formation* classes. However, the urban settlement in the center of the scene seems to be not fully captured. The model produces smaller patches of contiguously classified areas. When comparing this to the satellite image, the model seems to underestimate the size of the urban area and misclassifies parts of it as *Pasture* and *Annual and Perennial Crop*. The model also doesn't fully capture the patches of *Savanna Formation* within the *Forest Formation* south of the urban area. Although, when comparing this to the satellite base map in Fig. 4 it is not clear if these areas are truly savanna instead of forest.

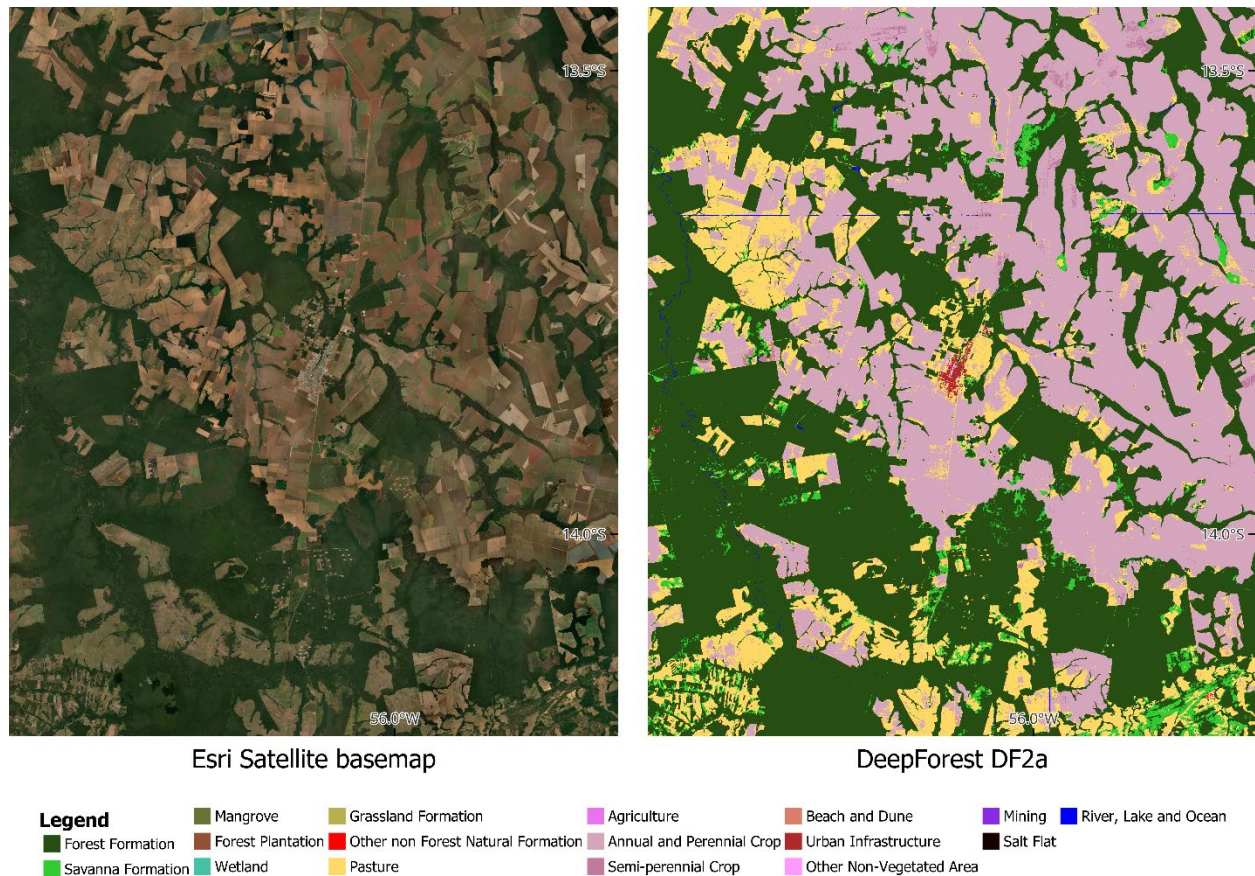


Fig. 4: The resulting land use and land cover map produced by the DF2a model (right) compared to the Esri Satellite basemap (left).

## 5 Discussion

The results, surprisingly, show some improvement in classification accuracies. However, this improvement is probably due to the underrepresented classes in the original dataset as well as class imbalance in the test dataset. Two major classes make up more than two thirds of the scene (*Forest Formation*: 32.4%; *Annual and Perennial Crop*: 37.2%) while the other widely distributed classes, *Savanna Formation* and *Pasture*, cover 18.1% and 10.1% of the scene, respectively. This imbalance is present in the whole Amazon basin, where *Forest Formation* covers 43.7% of the whole area. Together with the other dominating classes: *Pasture*, *Grassland*, and *Savanna Formation*. These four classes define over 80% of the Amazon basin. Although the models were also trained on this imbalanced data set, a weighting factor was added to the loss function to account for the imbalance.

Another factor contributing to the improvement in overall accuracies might be the change in the classification scheme and the used methodology to derive the classes in the MapBiomass project itself. It can be argued that the new scheme better represents landcover classes and is more accurate than previous collections and thus reduces the error produced by inaccurate labels. To further investigate these effects, we will train the models on larger, new datasets with the new label

collection and compare results to the present transfer study. The qualitative assessment with the produced land use and land cover maps shows consensus with the true classes when visually comparing these with a higher resolution satellite base map and, thus, highlight the potential of these models for automated classification.

## 6 Literature

- CHERIF, E., HELL, M. & BRANDMEIER, M., 2022: DeepForest: Novel Deep Learning Models for Land Use and Land Cover Classification Using Multi-Temporal and -Modal Sentinel Data of the Amazon Basin. *Remote Sensing*, **14**(19), 5000, <https://doi.org/10.3390/rs14195000>.
- ESA, 2023: SNAP – STEP. <https://step.esa.int/main/toolboxes/snap/>, letzter Zugriff 09.01.2023.
- ESRI, 2023: ArcGIS Pro | 2D, 3D & 4D GIS Mapping Software. <https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview>, letzter Zugriff 25.01.2023.
- FILIPPONI, F., 2019: Sentinel-1 GRD Preprocessing Workflow. *International Electronic Conference on Remote Sensing, Proceedings*, **18**(1), 11, <https://doi.org/10.3390/ECRS-3-06201>.
- MAPBIOMAS, 2022: MapBiomas Brasil. <https://mapbiomas.org/en>, letzter Zugriff am 17.03.2022.
- SHI, X., CHEN, Z., WANG, H., YEUNG, D.-Y., WONG, W. & WOO, W., 2015: Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in neural information processing systems*, **28**.

# Monitoring Crop Seasonal Development under Biochar Treatment with Unmanned Aerial Vehicle (UAV)-based Multispectral Imagery

SAJA SALATTNA<sup>1</sup>, JULIANE BENDIG<sup>1</sup>, CHRISTINA KUCHENDORF<sup>1</sup>,  
CHRISTOPH JEDMOWSKI<sup>1</sup> & UWE RASCHER<sup>1</sup>

*Abstract: Remote sensing data plays a crucial role in crop growth monitoring by characterizing the spatiotemporal variability of crop traits. In the context of climate-smart agriculture, the impact of biochar application in agroecosystems has received significant attention in recent years. However, despite the numerous studies examining the potential of biochar-facilitated fertilization on harvested crop yields and above-ground biomass production, monitoring of the impact of biochar treatments by remote sensing approaches on crop dynamics during the growing season has not been well addressed. Therefore, the main objective of this research was to provide a first case study to evaluate the potential of optical remote sensing to investigate the impact of biochar on crop growth. We used seasonal multispectral data acquired from an unmanned aerial vehicle (UAV) with high spatial resolution. In our case study, we used an off-the-shelf DJI Phantom 4 multispectral UAV system to monitor the seasonal development of spelt (*Triticum aestivum* ssp. *spelta* L.) in a biochar enrichment experiment. We here propose a straightforward data processing workflow based on an empirical line method to convert raw UAV data to normalized and comparable reflectance maps, which then were the bases for calculating various visible/near-infrared (VIS/NIR) vegetation indices. OSAVI was the best resolving index in relation to the actual yield, indicating equal to better spelt development over the biochar plots with a fertilizer level comparable to half and full conventional amount, compared to conventionally high fertilized controls.*

## 1 Introduction

Global plant production faces the main challenge of sustainability under the constraint of the rapidly increasing population and the gradual depletion of natural resources (FRÓNA et al. 2019). In this context, reliable and timely information about the crop sown area becomes essential in developing management strategies in precision agriculture (RAMTEKE et al. 2020). Precision agriculture is a site-specific crop management approach that gathers, processes, and analyses spatial and temporal variability in soil, field, and crop parameters to support management decisions (SISHODIA et al. 2020). Remote sensing plays a crucial role in precision agriculture as it covers large areas rapidly and repeatedly, thus characterizing spatial and temporal crop traits. In addition to providing a synoptic view, remote sensing can provide structural information about vegetation health. As the spectral reflectance of the crop canopy changes with phenology (growth), stage and crop health, it can be measured and monitored by multispectral sensors (RAMTEKE et al. 2020) .

---

<sup>1</sup> Institute of Bio- and Geosciences: Plant Sciences (IBG-2),  
Forschungszentrum Jülich GmbH, 52428 Jülich, Germany,  
E-Mail: [s.salattna, j.bendig, c.kuchendorf, c.jedmowski, u.rascher]@fz-juelich.de

Satellite images have been widely adopted for crop monitoring due to the images' often sufficient spatial, spectral, and temporal resolution (LI et al. 2022; BLICKENSDÖRFER et al. 2022). Furthermore, high-resolution unmanned aerial vehicle (UAV)-based imagery has been used recently to address the data gap in satellite products suffering from cloud coverage and to address questions requiring centimeter-level spatial resolution (YAO et al. 2019).

In the context of mitigating climate change and adapting to the loss of soil fertility, there is a particular interest in assessing the impact of biochar-facilitated fertilizer application in agroecosystems (JEFFERY et al. 2011). Biochar is a black carbon-rich product obtained through the pyrolysis of various biomass feedstocks to intentionally amend the soil. Several studies have indicated that biochar application to soil can (i) improve soil fertility, (ii) enhance nutrient availability, (iii) increase soil water holding capacity and correlated growth and yield, (iv) increase microbial population and activities, and (v) reduce greenhouse gas emissions through carbon sequestration (TRUPIANO et al. 2017; BIEDERMAN & HARPOLE 2013). A commonly reported impact of biochar is the improvement of above-ground biomass productivity. Nevertheless, previous studies found that biochar could affect crop development only during the green-up phase, while the harvested crop yield could remain unaffected (HEIDARIAN DEHKORDI et al. 2020b).

Despite the growing literature investigating the effect of biochar on total crop productivity, the biochar effects on crop development over the growing season have received much less attention. Therefore, the spatio-temporal information provided by the UAV images can reveal variability in crop performance due to the presence of biochar within the agricultural soils. So far, HEIDARIAN DEHKORDI et al. (2020a) (2020b) have investigated the advanced perspectives of UAV (10 channel MicaSense RedEdge-M) remote sensing in combination with in-situ ground measurements of crop traits for monitoring the crops dynamic affected by century-old biochar enrichment in precision agriculture. To date, short-term biochar effects on plant systems using high-resolution UAV-based data at the canopy level have yet to be addressed.

The primary objective of this manuscript is to describe the ability to use the small-scale, high spatial-resolution UAV system to detect the differences between biochar fertilized fields and control fields and to investigate at which development stage those differences can be quantified or detected. Due to the high fertilization standard of the high-quality field, the expected impact of biochar-facilitated fertilization on crop growth may be low but possibly detectable using high spatial-resolution UAV data. The available imagery source was visible-near infrared multispectral data collected over one crop growth season. This study proposes a data processing workflow, from raw imagery to creating a reflectance map and extracting vegetation indices (VIs). These indices are used to answer the following question: Is it possible to detect the short-term impact of biochar treatments on spelt (*Triticum aestivum ssp. spelta* L.) crop performance at the canopy level using UAV-based crop traits?

## 2 Material and Methods

### 2.1 Site description

The study area was a farmer's field of approximately 21 hectares cultivated with spelt, located in Germany (latitude: 50° 58'55.1"N, and longitude: 6°25'50.4"E ) (Figure 1). Spelt seeds were sown

in December 2022. The Muencheberg soil quality rating (SQR) of the area is  $>85$ , Colluvisol over Chernozem-parabraunerde. Spelt development was monitored during the experiment from the Booting stage to the senescence. The experimental treatment consisted of three biochar-enriched stripes with a diameter of 6 m along the field's length. Each stripe was treated with 1 ton/ha of terra preta biochar and 180 kg nitrogen/ha, 90 kg N/ha, and 40 kg N/ha for stripes T1, T2, and T3, respectively. Apart from these biochar-enriched stripes, the experimental field was exposed to homogeneous agricultural practices, and the soil was treated with organic minerals of 180 kg N/ha.

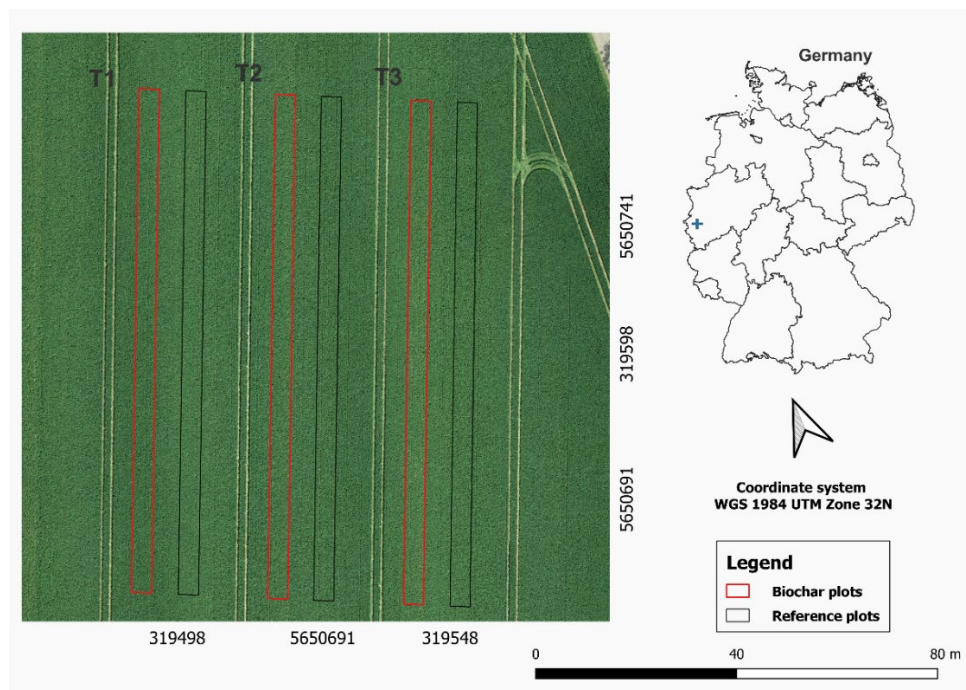


Fig. 1: Map of the experimental pairs (reference vs biochar plots) in the spelt field in Ameln-Titz, Germany. Background image corresponds to the Red-Green-Blue (RGB) orthomosaic captured by UAV on 02 June 2022

Three  $3 \times 100$  m plots (red stripes in Fig. 1) were selected inside the biochar-enriched stripes, at least 1.5 m from the border of each stripe, to avoid gradient effects at the edge of the treatments. Subsequently, three plots of  $3 \times 100$  m (black stripes in Fig. 1) were selected as reference soil areas within a distance of approximately 3 m from the biochar-enriched stripes to ensure comparable soil properties while preventing effects from mixing of biochar and reference soil particles. A set of ground control points (GCPs), consisting of five  $0.3 \times 0.3$  m targets, were placed in the field for geo-referencing UAV images. GCP coordinates were measured with a real-time kinematic (RTK) global navigation satellite system (GNSS) (Hiper VR, Topcon Positioning Systems, Inc., Tokyo, Japan) with an overall accuracy of 0.01 m.

## 2.2 Unmanned Aerial Vehicle (UAV) data acquisition

A multispectral remote sensing dataset was acquired during the 2022 growing season using the multispectral camera mounted on the DJI Phantom 4 Multispectral (P4M). The P4M camera has six bands, including one RGB sensor and five monochrome sensors capturing blue (450 nm central



wavelength), green (560 nm), red (650 nm), red-edge (730 nm), and near-infrared (840 nm). Each spectral sensor has a global shutter and 1600×1300 pixels image resolution. In addition, the P4M is equipped with a band-by-band incident light sensor, which allows the irradiance measurement for each band during flight. The UAV was flown at 50 m above ground level (AGL) with an airspeed of 4.25 ms<sup>-1</sup>. The forward and side image overlaps were 75% and 65%, resulting in a ground sampling distance (GSD) of 0.035 m. During data collection, six flights were performed between May and July 2022 in clear sky conditions and between 12:00 to 15:00 h local time. For each flight, a set of Lambertian reference panels with known hemispherical-conical reflectance factors was placed next to the field and recorded at flight altitude to enable the generation of reflectance maps in the postprocessing.

## 2.3 UAV-data preprocessing

### 2.3.1 Processing workflow

The data processing workflow, including data processing and vegetation indices extraction, is illustrated in Figure 2. First, the individual images were corrected for vignetting effect (section 2.3.2). Then, raw digital numbers (DNs) were corrected for sensor gain and exposure time yielding normalized-DN values (section 2.3.3). These two steps were performed on the individual images using an in-house developed python code. The corrected images were processed using the Metashape processing workflow (Agisoft LLC, St. Petersburg, Russia) that included GCPs for georeferencing orthomosaics as final products. The empirical line method (ELM) was applied for atmospheric correction and to create reflectance maps using a python code (section 2.3.4). Several vegetation indices were then calculated from the reflectance maps using the raster calculator tool in QGIS (section 2.4). In addition, a mask was applied to the vegetation indices maps in QGIS to exclude soil pixels before conducting the statistical tests to examine the difference between the reference and biochar plots (2.5).

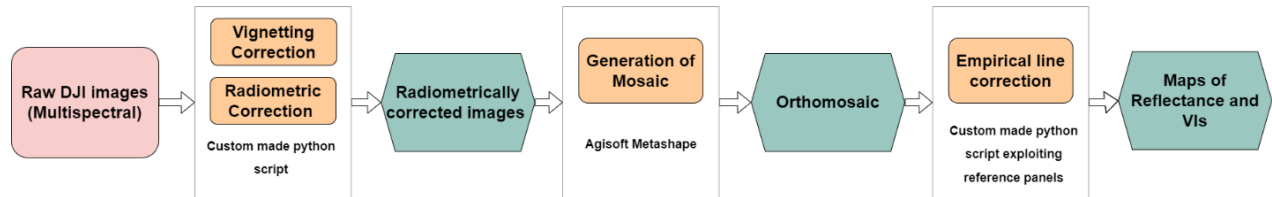


Fig. 2: Overall multispectral UAV-image processing workflow

### 2.3.2 Vignetting correction

Vignetting is the effect of the radial falloff in pixel values, which results in darker areas near the edges of images (Goldman 2010). The vignetting effect,  $V(x, y)$ , is typically modeled as a high-order polynomial (Equation 2). It assumes zero vignetting,  $V(x_v, y_v) = 1$ , at the image center:

$$r = \sqrt{(x - x_v)^2 + (y - y_v)^2} \quad (1)$$

$$V(x, y) = 1 + \alpha_1 r + \alpha_2 r^2 + \alpha_3 r^3 + \alpha_4 r^4 + \alpha_5 r^5 + \alpha_6 r^6 \quad (2)$$

Where  $r$  is the distance of the pixel  $(x, y)$  from the center of the vignette  $(x_v, y_v)$ , which can be obtained by (Equation 2), matrix  $\alpha$  shows the polynomial coefficients for the vignetting correction, which can be found in the EXIF/XMP metadata of the images.

The corrected image intensity ( $I_{corrected}$ ) is calculated from the original image intensity  $I(x, y)$  using Equation (3):

$$I_{corrected}(x, y) = [I(x, y) - BL(x, y)] \times V(x, y) \quad (3)$$

Where  $BL(x, y)$  is the normalized black level value, which can be found in EXIF/XMP metadata.

### 2.3.3 Radiometric correction

The radiometric correction is a crucial step in the processing flow. When an image is captured, raw data are stored as DNs, representing the radiant energy received by each pixel. During radiometric calibration, DNs are converted to radiance and usually later top of canopy reflectance to enable comparisons between datasets acquired under differing light conditions or with different sensors (MINAŘÍK et al. 2019). Radiance in absolute physical units cannot be calculated for DJI multispectral images as the sensor is not radiometrically calibrated by the manufacturer. Instead, normalized DNs are calculated using Equation 4, as described in Multispectral Image Processing Guide provided by DJI (DJI 2020):

$$DN_{corrected} = \frac{I_{corrected} \times 1e^6}{sensor\ gain \times camera\ exposure\ time} \quad (4)$$

where sensor gain and the camera exposure time can be found in XMP/EXIF metadata of each image, and  $1e^6$  is used to convert the exposure time from microseconds to seconds.

### 2.3.4 Mosaic generation

The individual images, corrected as described in sections 2.3.2 and 2.3.3, were imported into Agisoft Metashape to automate the image post-processing, recognize the characteristics of the used multispectral camera, and make the necessary geometric corrections (e.g., lens distortion). As the next step, the GPS metadata of the GCPs were used for geo-positioning and photogrammetric reconstruction, and finally, the software generated the orthomosaic and digital surface model.

### 2.3.5 Reflectance

ELM is commonly used for atmospheric correction and for generating surface reflectance data. The method assumes a linear relationship between DNs or radiance values in an image and surface reflectance (SMITH & MILTON 1999). Therefore, one or more reflectance calibration panels of known reflectance and Lambertian properties must be visible in the aerial imagery to apply the method. The UAV data for this study was acquired simultaneously with nine standard reflectance panels (2%, 3%, 4%, 5%, 6%, 12%, 24%, 40% and 63%). First, the reference panel ROIs were manually extracted from the central part of each panel in the orthomosaic, for every flight, in QGIS. Then, the reference panel measurements were obtained band by band using a python script,

due to a displacement between the bands in the DJI P4 multispectral imageries. A mask was also applied for each orthomosaic within the python script to mask out the saturated panels for each spectral band. Then, ELM was applied to the orthomosaic to create a reflectance map as described in (CHAKHVASHVILI et al. 2021).

### 2.3.6 Calculation of Multispectral vegetation indices

Vegetation indices allow monitoring spectral changes related to, e.g., crop structure, above-ground biomass, crop health, and weed presence (CUARAN & LEON 2021). The normalized difference vegetation index (NDVI) is a good indicator for green biomass, with the well-known shortcomings of saturating in closed canopies and being influenced by soil reflectance. The enhanced vegetation index (EVI) was proposed to address this issue by adjusting for atmospheric conditions and canopy background noise using a blue band. The normalized difference red edge index (NDRE) was identified as a good proxy of nitrogen concentration. The chlorophyll index red (CI-red) and the simplified canopy chlorophyll content index (s-CCCI) are often used for evaluating canopy chlorophyll and nitrogen content. The optimized soil-adjusted vegetation index (OSAVI) and chlorophyll vegetation index (CVI) are indicative of leaf chlorophyll content. Formulae and references are displayed in Tab. 1.

Tab. 1: Multispectral vegetation indices calculated in this study

Index		Formulae	Reference
Normalized Difference Vegetation Index	NDVI	$(\text{NIR} - \text{Red}) / (\text{NIR} + \text{Red})$	(BUSCHMANN & NAGEL 1993)
Normalized Difference Red Edge	NDRE	$(\text{NIR} - \text{RE}) / (\text{NIR} + \text{RedEdge})$	(GITELSON & MERZLYAK 1994)
Enhance Vegetation Index	EVI	$2.5 * [(\text{NIR} - \text{Red}) / (\text{NIR} + 6 * \text{Red} - 7.1 * \text{Blue} + 1)]$	(ZHANG et al. 2006)
Optimized soil adjusted VI	OSAVI	$1.16 (\text{NIR} - \text{Red}) / (\text{NIR} + \text{Red} + 1.16)$	(LIMA-CUETO et al. 2019)
Chlorophyll vegetation index	CVI	$(\text{NIR}/\text{Green}) * (\text{Red}/\text{Green})$	(VINCINI et al. 2008)
Chlorophyll index red	CI-red	$(\text{NIR}/\text{Red}) - 1$	(GITELSON et al. 2003)
Simplified canopy chlorophyll content index	s-CCCI	$\text{NDRE} / \text{NDVI}$	(FITZGERALD et al. 2006)

## 2.4 Statistical Analysis

Analysis of Variance (ANOVA) and Tukey-Kramer post hoc test (NANDA et al. 2021) were used to determine whether the contrasts between the mappings of reference and biochar plots were statistically significant. As expected, a significant difference was found at a very high level due to the high number of samples (n) and, therefore, low variance. Since this is inconclusive, Cohen's D as a standardized mean difference (AARTS et al. 2013) was used to estimate the distance between means as effect size. Since the data distribution of pixel values can be fitted with a gaussian function, a comparison of means with a factor is possible. Therefore, peaks of the overlapping Gaussian functions can be distinguished, indicating that the underlying images (treatments) are from different datasets and differ to a certain degree. Cohen's D (as effect size) and the means  $\pm$  standard deviation (SD) are discussed in this manuscript as means of differentiation and possibly

magnitude or contrast in VI representation. The effect size is interpreted using the classification developed by COHEN (1988), which is presented in Tab. 2:

Tab. 2: Cohen's D effect size classification

Effect size (D)	Interpretation
$0,00 \leq D < 0,20$	Ignored
$0,20 \leq D < 0,50$	low
$0,50 \leq D < 0,80$	Medium
$0,80 \leq D < 1,00$	Strong
$1,00 \leq D$	Very Strong

### 3 Results and Discussion

Figure 3 exemplarily illustrates the time series for OSAVI. The data collection started in May when the spelt was between the booting phase and inflorescence emergence. The mean of the observed areas (T1-R3) relates to plant health through chlorophyll content, which is, during the vegetative phase, a sign of healthy, well-supplied, and growing plants. The highest values are expected just before flowering, and mean values decline towards senescence.

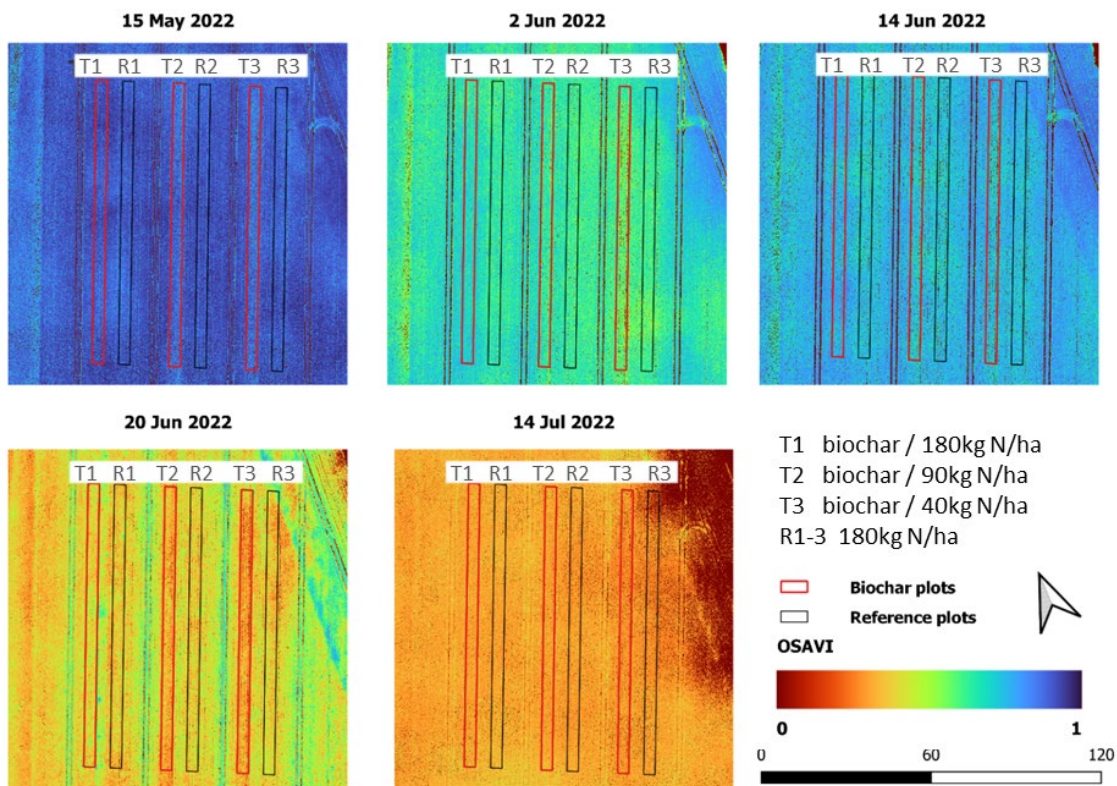


Fig. 3: Optimized soil adjusted vegetation index (OSAVI) time series; the boxes are the areas with biochar-fertilizer treatments as shown in Fig. 1. Colouring shows the seasonal dynamic in mean pixel values of this index

On the second measurement day (2<sup>nd</sup> June), OSAVI was lower compared to 15<sup>th</sup> May and rose again on 14<sup>th</sup> June, which is also apparent in the boxplot in Fig. 4. This response is likely related

to weather conditions (not shown), where the field experienced a dry and slightly too warm May with insufficient rain and some colder days with possible light stress towards 2<sup>nd</sup> June, followed by warm and moist enough conditions until 14<sup>th</sup> June. This response could be seen as a recovery on 14<sup>th</sup> June from a short stress period; from around 14<sup>th</sup> June on, there were the only short phases of optimal growing conditions in an otherwise too dry and warm growing season, though that fit well with spelt senescence timing. From 14<sup>th</sup> June onwards, OSAVI constantly decreased due to the onset of senescence until 14<sup>th</sup> July, one day before the harvest.

The mean OSAVI value of the biochar treatment 1 plot (red box) was higher than in the adjacent reference plot (grey box) (same order as in Fig. 3) at each acquisition date except on 14<sup>th</sup> July, when the mean OSAVI was 0.3 for both reference and biochar plots (Fig. 4). Cohen's D showed that the difference was nearly negligible on all the dates with a D value less than 0.2, except for 2<sup>nd</sup> and 29<sup>th</sup> June, which have a medium and low effect with  $D = 0.54$ , and  $0.37$ , respectively. For treatment 2, reported Cohen's D effects compared to the references except for 15<sup>th</sup> May and 29<sup>th</sup> June could be mostly ignored. In contrast, Cohen's D indicated a medium effect between treatment 3 and its adjacent reference on 15<sup>th</sup> May, 2<sup>nd</sup> and 29<sup>th</sup> June, and a medium to very strong effect toward treatment 1 with D values of  $.71$ ,  $1.18$ , and  $1.19$ , respectively. These numbers indicate that the spelt is denser and healthier in treatment 1 compared to the reference plots and treatment 2. Both performed better than treatment 3; an effect was visible by eye at harvest in July.

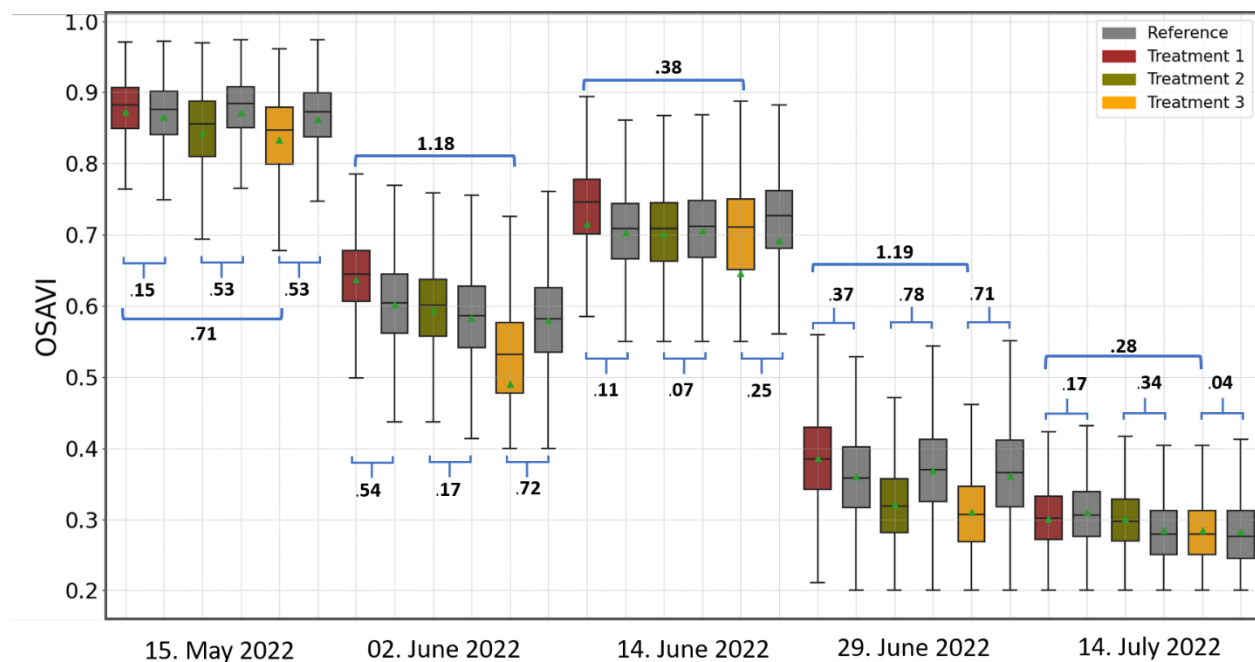


Fig. 4: Comparison of the optimized soil adjusted vegetation index (OSAVI) of the three biochar stripes versus the three reference stripes for each acquisition date. Whiskers show highest and lowest values respectively, dividing line is the median, green triangle the mean. Brackets indicate comparisons between treatments, with respective Cohen's D value

Graphs of other examined multispectral vegetation indices are not presented here but data can be found in Tab. 3 and 4. Biochar treatment effects became visible in all VIs. However, those effects were negligible on 15<sup>th</sup> May for nearly all the VIs except EVI, which showed a medium to strong

effect, and OSAVI indicated a Medium effect. On 2<sup>nd</sup>, 14<sup>th</sup>, and 29<sup>th</sup> June, NDVI, EVI, NDRE, and OSAVI exhibited medium to very strong impact with a D value ranging up to 0.5 and higher.

Tab. 3: Mean vegetation index values and maximum distance between means (%) for the biochar and reference plots during the growing season 2022

		15.05.2022	02.06.2022	14.06.2022	29.06.2022	14.07.2022
<b>NDVI</b>	total mean $\pm$ STD	0.91 $\pm$ 0.04	0.67 $\pm$ 0.08	0.79 $\pm$ 0.05	0.41 $\pm$ 0.09	0.3 $\pm$ 0.05
	max dist. btw. means [%]	1.70%	12.8%	3.13%	17.9%	11.2%
<b>EVI</b>	total mean $\pm$ STD	0.56 $\pm$ 0.1	0.63 $\pm$ 0.13	0.73 $\pm$ 0.1	0.27 $\pm$ 0.07	0.29 $\pm$ 0.05
	max diff. btw. means [%]	16.7%	24.0%	6.58%	22.0%	4.95%
<b>OSAVI</b>	total mean $\pm$ STD	0.86 $\pm$ 0.06	0.58 $\pm$ 0.1	0.69 $\pm$ 0.13	0.35 $\pm$ 0.07	0.29 $\pm$ 0.05
	max diff. btw. means [%]	4.48%	22.9%	9.67%	19.7%	5.15%
<b>NDRE</b>	total mean $\pm$ STD	0.39 $\pm$ 0.05	0.17 $\pm$ 0.05	0.2 $\pm$ 0.06	0.22 $\pm$ 0.04	0.16 $\pm$ 0.04
	max diff. btw. means [%]	7.14%	24.7%	20.9%	9.83%	6.58%
<b>Clred</b>	total mean $\pm$ STD	25.3 $\pm$ 9.49	4.23 $\pm$ 1.47	7.97 $\pm$ 3.17	1.43 $\pm$ 0.57	0.96 $\pm$ 0.29
	max diff. btw. means [%]	14.7%	33.0%	17.1%	29.4%	15.8%
<b>CVI</b>	total mean $\pm$ STD	6.88 $\pm$ 2.39	3.34 $\pm$ 1.01	9.01 $\pm$ 1.97	7.51 $\pm$ 2.25	4.28 $\pm$ 1.18
	max diff. btw. means [%]	15.9%	12.6%	14.8%	5.54%	15.9%
<b>s-CCCI</b>	total mean $\pm$ STD	0.42 $\pm$ 0.05	0.25 $\pm$ 0.07	0.26 $\pm$ 0.06	0.54 $\pm$ 0.11	0.53 $\pm$ 0.13
	max diff. btw. means [%]	8.45%	16.5%	17.5%	13.7%	6.99%

Overall, the findings correlate well with fertilization management, where treatment 1 received the same amount of conventional nitrogen fertilizer as was used for the reference under standard cultivation. Treatment 2 received half the amount of fertilizer, and treatment 3 received no additional fertilization except the replacement of formerly depleted N on the field, resulting in 22% less N than the references. As for the crop yield, the farmer reported that treatment 1 was about 5% higher than the reference yield, and treatment 2 was like the reference despite having received half the fertilization. However, treatment 3 yield was 20% less than the references (direct communication).

All examined vegetation indices showed negligible to low difference between biochar and reference plots on the 14<sup>th</sup> July. OSAVI and EVI performing best in the vegetative phase. CVI showed overall large variances and is likely not recommended for detecting the effects of biochar in a homogeneous grassy cereal canopy. For comparisons between VIs, the ability to resolve differences between treatments, expressed by a wide range of values, and resolution, relating to the change of mean over the whole season should be considered. It should be noted that the approach of effect-size-based evaluation, essentially comparable to an arbitrary calibration of colour values (via Cohen's D), appears to work well in this case and correlates with the actual productivity and observations.

Tab. 4: The magnitude of difference/distance between means by Cohen's D of the vegetation indices comparing the biochar-treated and reference plots; strong and very strong effects **bold (D > .80)**; T1-T3, the overall largest difference as expected between highly and lower-than-conventional fertilization, measured by final yield as well as most indices, **highlighted in grey**. Not all comparisons (15 overall) shown. Underscored values correlate with the maximum distance ([%], see table 2) for the respective index on that acquisition

Date/ Index		15 May 2022	02 Jun 2022	14 Jun 2022	29 Jun 2022	14 Jul 2022
NDVI	T1 - T2	0.05	0.62	0.54	0.77	0.09
	T1 - T3	0.17	<b>1.44</b>	0.46	<b>0.98</b>	0.41
	T2- T3	0.21	<b>0.84</b>	0.09	0.19	0.32
	T1 - R1	0.02	0.53	0.52	0.26	0.01
	T2 - R2	0.2	0.2	0.1	0.64	0.1
	T3 - R3	0.36	0.64	0.26	0.77	0.29
EVI	T1 - T2	<b>0.96</b>	0.63	0.49	<b>0.97</b>	0.13
	T1 - T3	<b>1.17</b>	<b>1.48</b>	0.32	<b>1.06</b>	0.26
	T2- T3	0.19	0.76	0.16	0.11	0.14
	T1 - R1	0.21	0.66	0.48	0.4	0.02
	T2 - R2	<b>0.86</b>	0.25	0.09	0.74	0.11
	T3 - R3	0.73	0.42	0.11	<b>0.81</b>	0.02
OSAVI	T1 - T2	0.55	0.67	0.12	<b>1.02</b>	0.01
	T1 - T3	0.71	<b>1.18</b>	0.38	<b>1.19</b>	0.28
	T2 - T3	0.16	<b>0.84</b>	0.35	0.2	0.34
	T1 - R1	0.15	0.54	0.11	0.37	0.17
	T2 - R2	0.53	0.17	0.07	0.78	0.34
	T3 - R3	0.51	0.72	0.25	0.71	0.04
NDRE	T1 - T2	0.25	0.72	<b>0.81</b>	0.04	0.08
	T1 - T3	0.58	<b>1.01</b>	<b>0.91</b>	0.23	0.1
	T2 - T3	0.29	0.28	0.09	0.26	0.02
	T1 - R1	0.11	0.3	0.29	0.1	0.08
	T2 - R2	0.24	0.24	0.4	0.27	0.05
	T3 - R3	0.38	0.62	0.51	0.15	0.07
Clred	T1 - T2	0.09	0.58	0.52	0.73	0.08
	T1 - T3	0.15	<b>1.31</b>	0.42	<b>0.92</b>	0.31
	T2- T3	0.26	0.78	0.11	0.18	0.22
	T1 - R1	0.02	0.53	0.48	0.27	0.03
	T2 - R2	0.19	0.18	0.12	0.6	0.08
	T3 - R3	0.36	0.6	0.29	0.73	0.22
CVI	T1 - T2	0.33	0.19	0.62	0.13	0.06
	T1 - T3	0.3	0.07	0.67	0.19	0.42
	T2 - T3	0.03	0.11	0.06	0.06	0.48
	T1 - R1	0.06	0.26	0.08	0.19	0.2
	T2 - R2	0.18	0.33	0.57	0.06	0.07
	T3 - R3	0.1	0.17	0.4	0.14	0.05
s-CCCI	T1 - T2	0.44	0.66	0.64	0.78	0.05
	T1 - T3	0.73	0.68	<b>0.83</b>	0.76	0.22
	T2 - T3	0.27	0.01	0.16	0	0.17
	T1 - R1	0.13	0.14	0.09	0.37	0.11
	T2 - R2	0.3	0.38	0.4	0.34	0.01
	T3 - R3	0.42	0.53	0.46	0.65	0.06

However, this approach may not be generalizable since there are pitfalls in the statistical applicability, which may not have appeared here due to high homogeneity, but which nevertheless must be considered. An example is the high number of degrees of freedom with the extremely high

n. It is not necessarily a true high n since other influences besides fertilization cannot be ruled out completely. In an extreme case, one could regard this experiment as one random sample with  $n=6$  and 5 degrees of freedom. This observation could be alleviated by, e.g., bootstrapping and increasing the number of observations artificially, which allows for a post-experiment randomization to see if the initial observations still hold. Such an analysis would enable analysing how small an experimental area could be to rule out side effects while still being valid, which is advisable for future experiments.

In conclusion, differentiating short-term biochar treatments via Cohen's D works for VIs derived from multispectral, high spatial resolution UAV images. The biochar treatment showed the most indicative impact on spelt crop during the inflorescence emergence. In direct comparisons, vegetation indices, especially OSAVI, enable evaluating plant status and possibly forecasting yield locally, which may help to optimize soil pre-treatment or fertilization. However, comparisons with ground truth data from yield and scoring plants throughout the season should be performed to solidify the findings. In addition, the minimum plot size for differentiation of biochar-facilitated fertilization should be determined in the future.

## 4 Acknowledgment

This work was supported by the BMBF programs Palestinian-German Science Bridge (PGSB), FKZ 01DH16027 and the 'Strukturwandel-Projekt BioökonomieREVIER (FKZ 031B0918A). The authors would like to acknowledge and thank the following individuals for their expertise and assistance throughout the study: Onno Muller, Andreas Dering, Hendrik Poorter, and Hanno Scharr.

## 5 References

- AARTS, S., VAN DEN AKKER, M. & WINKENS, B., 2013: The importance of effect sizes (1381-4788), 61-64, <https://oi.org/10.3109/13814788.2013.818655>.
- BLICKENDÖRFER, L., SCHWIEDER, M., PFLUGMACHER, D., NENDEL, C., ERASMI, S. & HOSTERT, P., 2022: Mapping of crop types and crop sequences with combined time series of Sentinel-1, Sentinel-2 and Landsat 8 data for Germany. *Remote Sensing of Environment*, **269**, 112831, <https://oi.org/10.1016/j.rse.2021.112831>.
- BUSCHMANN, C. & NAGEL, E., 1993: In vivo spectroscopy and internal optics of leaves as basis for remote sensing of vegetation. *International Journal of Remote Sensing*, **14**(4), 711-722, <https://oi.org/10.1080/01431169308904370>.
- CHAKHVASHVILI, E., SIEGMANN, B., BENDIG, J. & RASCHER, U., 2021: Comparison of Reflectance Calibration Workflows for a UAV-Mounted Multi-Camera Array System, 8225-8228, <https://oi.org/10.1109/IGARSS47720.2021.9555143>.
- COHEN, J., 1988: *Statistical Power Analysis for the Behavioral Sciences*. 2<sup>nd</sup> ed. New York: NY: Routledge Academic.
- CUARAN, J. & LEON, J., 2021: Crop Monitoring using Unmanned Aerial Vehicles: A Review. *AG (Of)*, <https://oi.org/10.18805/ag.R-180>.



- DJI, 2020: P4 Multispectral Image Processing Guide. [https://dl.djicdn.com/downloads/p4-multispectral/20200717/P4\\_Multispectral\\_Image\\_Processing\\_Guide\\_EN.pdf](https://dl.djicdn.com/downloads/p4-multispectral/20200717/P4_Multispectral_Image_Processing_Guide_EN.pdf), last access 05.02.2023.
- FITZGERALD, G. J., RODRIGUEZ, D., CHRISTENSEN, L. K., BELFORD, R., SADRAS, V. O. & CLARKE, T. R., 2006: Spectral and thermal sensing for nitrogen and water status in rainfed and irrigated wheat environments. *Precision Agric*, **7**(4), 233-248, <https://oi.org/10.1007/s11119-006-9011-z>.
- FRÓNA, D., SZENDERÁK, J. & HARANGI-RÁKOS, M., 2019: The Challenge of Feeding the World **11**(20), 5816, <https://oi.org/https://doi.org/10.3390/su11205816>.
- 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, [https://oi.org/10.1016/S0176-1617\(11\)81633-0](https://oi.org/10.1016/S0176-1617(11)81633-0).
- GITELSON, A. A., GRITZ, Y. & MERZLYAK, M. N., 2003: Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *Journal of Plant Physiology*, **160** (3), 271-282, <https://oi.org/10.1078/0176-1617-00887>.
- GOLDMAN, D. B., 2010: Vignette and exposure calibration and compensation. *IEEE transactions on pattern analysis and machine intelligence*, **32**(12), 2276-2288, <https://oi.org/10.1109/TPAMI.2010.55>.
- HEIDARIAN DEHKORDI, R., BURGEON, V., FOCHE, J., PLACENCIA GOMEZ, E., CORNELIS, J.-T., NGUYEN, F. et al., 2020a: Using UAV Collected RGB and Multispectral Images to Evaluate Winter Wheat Performance across a Site Characterized by Century-Old Biochar Patches in Belgium. *Remote Sensing*, **12**(15), 2504, <https://oi.org/10.3390/rs12152504>.
- HEIDARIAN DEHKORDI, R., DENIS, A., FOCHE, J., BURGEON, V., CORNELIS, J. THOMAS, TYCHON, B. et al., 2020b: Remotely-sensed assessment of the impact of century-old biochar on chicory crop growth using high-resolution UAV-based imagery. *International Journal of Applied Earth Observation and Geoinformation*, **91**, 102147, <https://oi.org/10.1016/j.jag.2020.102147>.
- JEFFERY, S., VERHEIJEN, F., VAN DER VELDE, M. & BASTOS, A. C., 2011: A quantitative review of the effects of biochar application to soils on crop productivity using meta-analysis. *Agriculture, Ecosystems & Environment*, **144**(1), 175-187, <https://oi.org/10.1016/j.agee.2011.08.015>.
- LI, M., SHAMSHIRI, R. R., WELTZIEN, C. & SCHIRRMANN, M., 2022: Crop Monitoring Using Sentinel-2 and UAV Multispectral Imagery: A Comparison Case Study in Northeastern Germany. *Remote Sensing*, **14**(17), 4426, <https://oi.org/10.3390/rs14174426>.
- LIMA-CUETO, F. J., BLANCO-SEPÚLVEDA, R., GÓMEZ-MORENO, M. L. & GALACHO-JIMÉNEZ, F. B., 2019: Using Vegetation Indices and a UAV Imaging Platform to Quantify the Density of Vegetation Ground Cover in Olive Groves (*Olea Europaea* L.) in Southern Spain. *Remote Sensing*, **11**(21), 2564, <https://oi.org/10.3390/rs11212564>.
- MINAŘÍK, R., LANGHAMMER, J. & HANUŠ, J., 2019: Radiometric and Atmospheric Corrections of Multispectral  $\mu$ MCA Camera for UAV Spectroscopy. *Remote Sensing*, **11**(20), 2428, <https://oi.org/10.3390/rs11202428>.

- NANDA, A., MOHAPATRA, B. BHUSAN, MAHAPATRA, A. PRASADA KUMAR, MAHAPATRA, A. PRASAD KUMAR & MAHAPATRA, A. PRASAD KUMAR, 2021: Multiple comparison test by Tukey's honestly significant difference (HSD): Do the confident level control type I error. *Int. J. Stat. Appl. Math.*, **6**(1), 59-65, <https://oi.org/10.22271/math.2021.v6.i1a.636>.
- RAMTEKE, I. K., RAJANKAR, P. B., OBI REDDY, G. P., KOLTE, D. M. & SEN, T. K., 2020: Optical remote sensing applications in crop mapping and acreage estimation: A review **2**(4).
- SISHODIA, R. P., RAY, R. L. & SINGH, S. K., 2020: Applications of Remote Sensing in Precision Agriculture: A Review. *Remote Sensing*, **12**(19), 3136, <https://oi.org/10.3390/rs12193136>.
- SMITH, G. M. & MILTON, E. J., 1999: The use of the empirical line method to calibrate remotely sensed data to reflectance. *International Journal of Remote Sensing*, **20** (13), 2653-2662, <https://oi.org/10.1080/014311699211994>.
- TRUPIANO, D., COCOZZA, C., BARONTI, S., AMENDOLA, C., VACCARI, F. PRIMO, LUSTRATO, G. et al., 2017: The Effects of Biochar and Its Combination with Compost on Lettuce (*Lactuca sativa* L.) Growth, Soil Properties, and Soil Microbial Activity and Abundance. In *International Journal of Agronomy*, 1-12, <https://oi.org/10.1155/2017/3158207>.
- VINCINI, M., FRAZZI, E. & D'ALESSIO, P., 2008: A broad-band leaf chlorophyll vegetation index at the canopy scale. *Precision Agric*, **9**(5), 303-319, <https://oi.org/10.1007/s11119-008-9075-z>.
- YAO, H., QIN, R. & CHEN, X., 2019: Unmanned Aerial Vehicle for Remote Sensing Applications - A review. *Remote Sensing*, **11** (12), 1443, <https://oi.org/10.3390/rs11121443>.
- ZHANG, X., FRIEDL, M. A. & SCHAAF, C. B., 2006: Global vegetation phenology from Moderate Resolution Imaging Spectroradiometer (MODIS): Evaluation of global patterns and comparison with in situ measurements. *J. Geophys. Res.*, **111**(G4), <https://oi.org/10.1029/2006JG000217>.

# A Two-Step Approach for the Acquisition of Individual Tree Outlines using Paid Crowdsourcing

DAVID COLLMAR<sup>1</sup>, VOLKER WALTER<sup>1</sup>, MICHAEL KÖLLE<sup>1</sup> & UWE SÖRGEL<sup>1</sup>

*Abstract: In this paper, we aim to investigate the acquisition of tree outlines from image data using paid crowdsourcing. A common approach for improving data quality in crowdsourcing is multiple acquisition followed by integration, which can be done, for example, by majority voting. Based on this, we present a two-step approach to obtain individual tree outlines from orthophotos. A first crowd campaign is used to segment and detect relevant areas. Here, the dataset is annotated by several crowdworkers with a subsequent integration by majority voting. This is followed by a second campaign in which the tree outlines from the previously selected relevant areas are collected by multiple crowdworkers by means of polygons. These polygons are then integrated into one single polygon per tree using a pixel-based voting approach. Depending on the threshold, an improvement in Intersection over Union (IoU) of up to 13 percentage points can be achieved, with a standard deviation reduced by up to 40%.*

## 1 Introduction

Due to the success of machine learning (ML), there is an enormous demand for training data (STONEBRAKER & REZIG 2019). While much effort is put into machine learning algorithms, only a fraction of the effort is spent on training data preparation (STONEBRAKER & REZIG 2019). Training data are needed in large quantities, and low-quality training data cannot be compensated even by the best ML algorithms (WHANG & LEE 2020). Therefore, high quality training data are essential and the collection of such data is an important research topic. A well-known approach for collecting training data, potentially at high quality, is crowdsourcing (WHANG & LEE 2020).

There are many publications that deal with the use of crowdsourcing for generation of training data. According to SARALIOGLUA & GUNGOR (2020), a majority of studies in the field of remote sensing use crowdsourcing for the acquisition of training data for image classification or data acquisition, and potential contributions through crowdsourcing “enable artificial intelligence to develop very quickly” (SARALIOGLUA & GUNGOR 2020).

In general, crowdsourcing comes in two types: Voluntary crowdsourcing and paid crowdsourcing. Voluntary crowdworkers are mostly motivated by intrinsic factors: No monetary or comparable compensations are expected, instead, crowdworkers are voluntarily participating in campaigns because it might be their hobby or general interest (HOSSAIN 2012; RYAN & DEZI 2000). Due to the difficulty of planning such campaigns, paid crowdsourcing can be an alternative. Here, extrinsic motivation is achieved through monetary compensation, which is one of the main motivations of users on paid crowdsourcing platforms (HOSSAIN 2012; PILZ & GEWALD 2013; YUEN et al. 2012).

---

<sup>1</sup> University of Stuttgart, Institute for Photogrammetry, Geschwister-Scholl-Str. 24D, 70174 Stuttgart, Germany, E-Mail: [David.Collmar, Volker.Walter, Michael.Koelle, Uwe.Soergel]@ifp.uni-stuttgart.de

Paid crowdsourcing can be used to generate training data in a fast and easy manner (SARALIOGLUA & GUNGOR 2020), at low cost (HIRTH et al. 2011) and with high quality of data – provided certain principles are followed (CHANDLER et al. 2013; CHEUNG et al. 2017). Quality improvements through task design may be included (ALLAHBAKSH et al. 2013; ZHANG et al. 2016), seeking “to guide the labelers to provide high quality labels” (ZHANG et al. 2016).

Crowdsourcing can be further enhanced by improvement of quality after data collection (ZHANG et al. 2016). This can be done by repeated labeling, where different workers are labeling the same dataset with subsequent integration. This approach is based on the theory of "Wisdom of Crowds", which follows the principle that “the output of the crowd can be greater than the sum of its parts” (CHANDLER et al. 2013). In addition to a potential increase in quality through multiple acquisitions (ZHANG et al. 2016), this also reduces the impact of so-called "satisficers", who are trying to maximize their earnings with minimal effort (CHANDLER et al. 2013) and thereby compromise quality. Paid crowdtasks can be published on platforms such as Amazon MTurk or Microworkers.com, which handle the recruitment and payment of workers. Crowdworkers registered on such platforms can freely select their job from those offered (HIRTH et al. 2011).

Crowdsourcing has been increasingly used in a wide variety of fields, including geospatial sciences: Both HADI et al. (2022) and SARALIOGLUA & GUNGOR (2022) use voluntary crowdsourcing for a crowd-based large-scale land classification using high-resolution satellite images, while PUTTINAOVARAT et al. (2022) apply it to assess flood damages. WALTER et al. (2020) extracted trees from 3D point clouds based on paid crowdsourcing, and WALTER et al. (2021) presented an approach to collect vehicles from point clouds by means of crowdsourcing using a two-level approach.

We investigate the potential of using the advantages of crowdsourcing for an efficient and cost-effective collection of high-quality tree polygons for later use as training data for machine learning in this paper. The rest of the paper is organized as follows: We first motivate and describe the developed methodology in sections 2 and 3. Section 4 presents the used dataset, while section 5 describes the performed experiments. Section 6 discusses the data analysis in detail, while section 7 ends this paper with the conclusion.

## 2 Preliminary Investigation

The success of crowd tasks depends strongly on task complexity: A too complicated task cannot be solved in a reliable way by means of crowdsourcing (ALLAHBAKSH et al. 2013), which is why the crowdworkers’ ability should be tested beforehand for tasks that might be challenging (CHANDLER et al. 2013). Since we are interested in high-quality annotations of individual tree outlines, which might be a rather complex issue, we followed this suggestion. Therefore, a test acquisition campaign was performed: We presented an image tile containing multiple trees of different size and shape (13 complete and 3 partial trees) to 25 crowdworkers on the crowdsourcing platform Microworkers.com. The crowdworkers were instructed to annotate all visible trees as precise as possible using individual polygons. The results are shown in Figure 1.

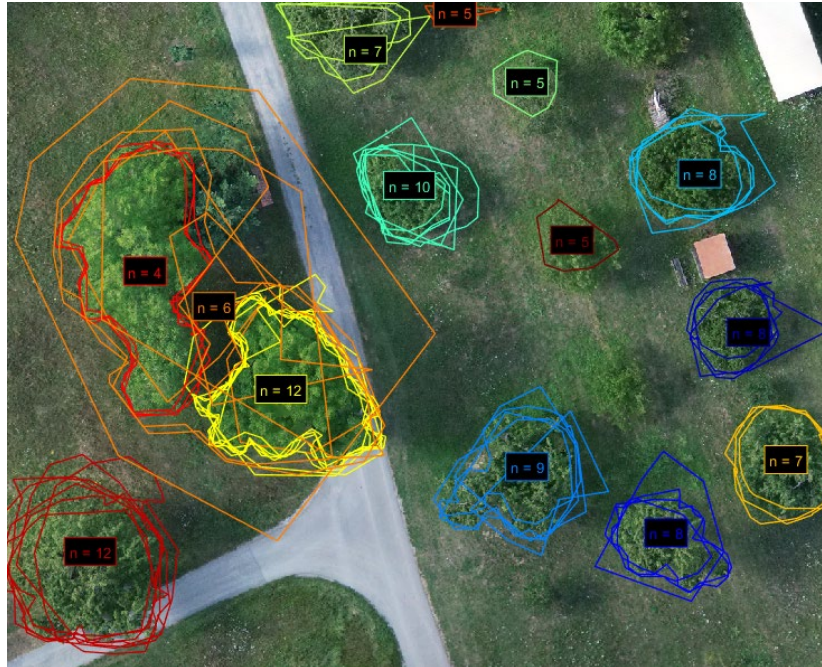


Fig. 1: Filtered crowd acquisitions on a sample image tile for 25 workers, color-coded by cluster

This task led to highly heterogeneous results. In total 151 trees were annotated, resulting in only approximately 6 annotations per crowdworker. After automatically removing invalid polygons and spam acquisitions, 114 polygons were left, leading to an effective 4.56 annotations per crowdworker of vastly different quality. The annotations were not only unprecise, which can be a common issue in crowdsourcing for various reasons (CHEUNG et al. 2017), but also incomplete (multiple trees not annotated at all), incorrect (two trees in central left part of the image annotated as one tree) and inhomogeneously distributed (different amount of annotations per tree). The reason for this high heterogeneity seems to lie in the task itself: Workers are effectively doing two steps at once (localization and annotation of trees via polygons), leading to a too complex and time-consuming microtask: Microtasks have a typical working time of some minutes and typically low salary (HIRTH et al. 2011). The results are not surprising and in line with the observations of FINNERTY et al. (2013): Having crowdworkers perform a dual-task can increase the cognitive demand of the crowdworker too much and therefore significantly decrease performance.

In order to reduce the complexity of the task and ensure homogeneity of results, we choose to split the dual-task into two separate steps. Both steps are designed to be processed by the crowd, keeping the manual work at a minimum. The first step serves to identify relevant areas using a grid approach, whereas the second step aims to capture the tree outlines with precise polygons. Those two separated steps both consist of a clear task definition and adequate granularity, making it possible to conduct quality control through task design (ALLAHBAKHSI et al. 2013; ZHANG et al. 2016). Both steps also include a simple preparatory test which workers must pass. We hope to prevent insufficient effort responding through this method, following the recommendation of CHEUNG et al. (2017).

### 3 Methodology

#### 3.1 Processing pipeline

The identification of relevant areas containing trees is realized by placing a grid consisting of square-shaped grid cells over the image data. Users are then asked to annotate each grid cell individually, having only the options to choose between “tree” and “no tree” – effectively performing a binary segmentation. This principle can be seen in Figure 2a. The shown 8×8 grid is chosen for demonstration purposes only. For real use cases, the cell size should be chosen much smaller, resulting in more grid elements (e.g. 50×50).

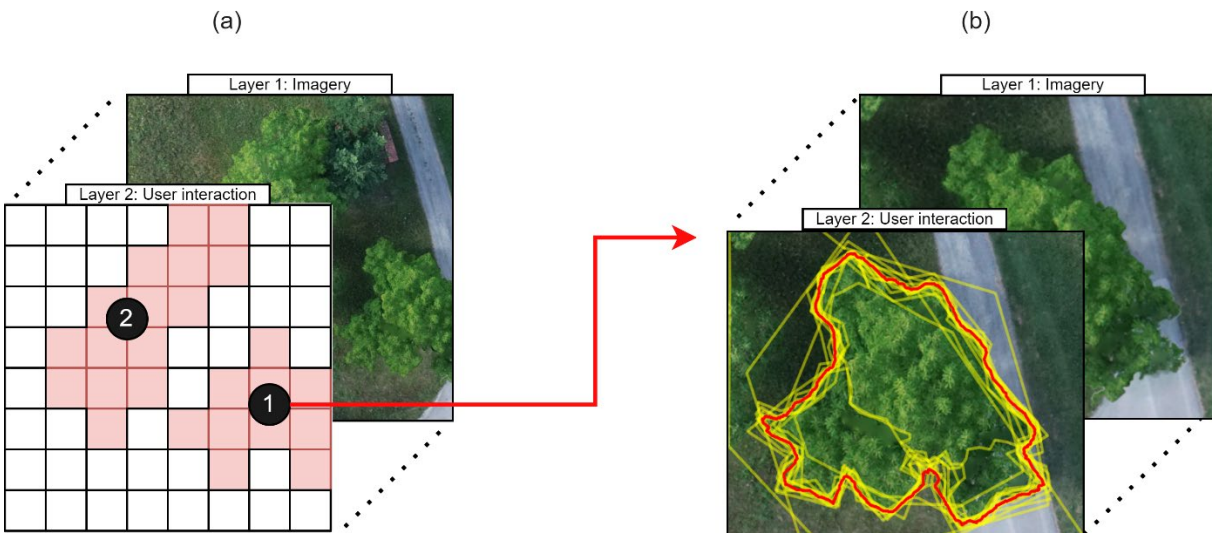


Fig. 2: Used two-step approach. (a) Step 1: Identification of relevant areas, (b) Step 2: Acquisition of precise tree outlines

Since we are using multiple labeling and are working on a binary classification, we can integrate our results using majority voting, a simple but effective method, as long as more than half of the workers provide correct results (ZHANG et al. 2016).

The integrated result is a downsampled segmentation map, which can be used for further processing. With this approach, we not only save costs and time, but also further simplify the task. Downsampling is possible since we are only interested in coarse tree positions and no pixel-precise segmentation is required for this approach. The degree of downsampling depends on the selected cell size.

Using the generated, downsampled segmentation map, tree clusters can easily be identified as closed areas. Each cluster can contain either a single tree or an area consisting of multiple trees. We are only interested in areas containing single trees, since we want to reduce complexity for the next task, following the suggestions of FINNERTY et al. (2013) to keep tasks simple. Therefore, a method for the distinction of areas containing only single trees from those containing multiple trees is required. Although there is no general solution for this, it is possible to draw conclusions about the number of trees in a selected area based on boundary conditions such as perimeter, area

or aspect ratio. In the example in Figure 2a, two clusters are found (indicated by numbers 1 and 2). In Figure 2, cluster 1 is taken for further demonstration of our approach.

Following the identification of relevant cluster areas, smaller image tiles containing only a single relevant cluster, therefore containing only individual trees, are cut out. These image tiles are then used as input for the second step, where precise tree outlines are annotated by means of polygons (see Figure 2b).

### 3.2 Integration of raw acquisitions

Multiple polygons of different quality are collected per tree, as can be seen in Figure 2b. Since only a single polygon per tree is desired as output, an integration operation must be performed. While the integration of classifications or even simple geometric shapes such as circles or rectangles can be easily solved, there is no trivial solution for more complex geometric shapes such as irregular polygons.

However, a fairly simple strategy can be used to integrate the polygons: A majority vote for each pixel of the RGB input data. Since users were annotating the polygons on pixel-level from the RGB data, we can convert all collected polygons to raster data. If the cell size of the raster representation is chosen to match the pixel size, there is only a small loss in information, such as the original positions of vertices. After the raster conversion, a pixelwise integration based on a binary vote can be performed: The number of different user labels for each raster cell is counted. If that number falls below a certain threshold, the pixel is omitted from the integrated raster shape, otherwise this pixel is included in the integrated raster shape.

This approach is shown in a very simplified form in Figure 3. For this figure, we chose the value 8 as the threshold for the binary voting, since this corresponds to an absolute majority for  $n=15$ , i.e., the number of crowdworkers who annotated one tree. As a result, any pixel with 8 or more acquisitions belongs to the integrated raster shape (Figure 3b). The integrated raster shape (Figure 3c) is then converted back into a polygon geometry, providing us with a tree outline polygon acquired through crowdsourcing.

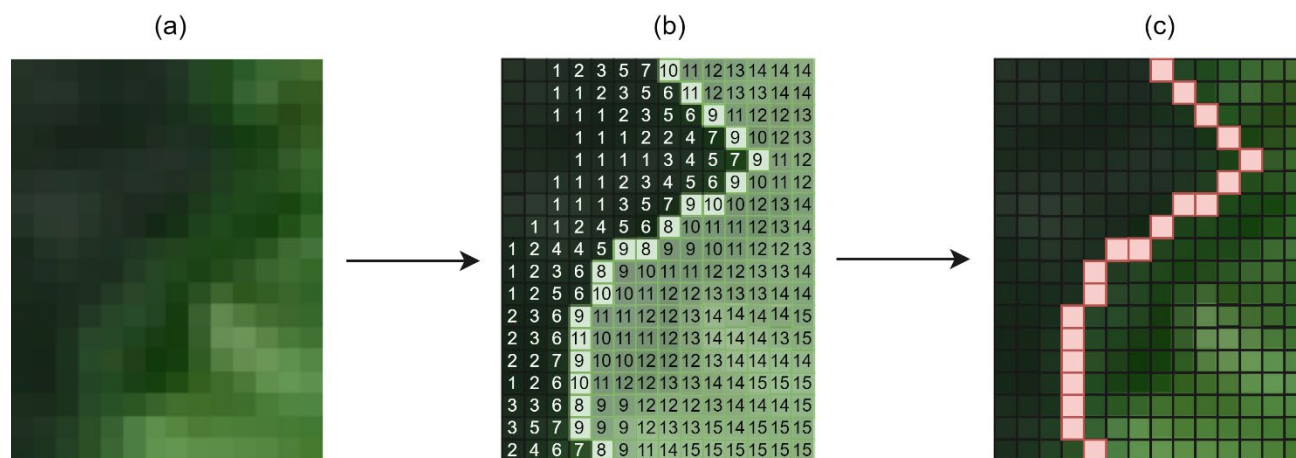


Fig. 3: Visualization of raster integration. (a) Section of image tile, (b) Majority vote on pixel basis with highlighted threshold areas, (c) Integrated polygon outline

## 4 Dataset

Our approach was tested for a designated test area near the town of Erligheim in southwestern Germany. The area mainly contains cherry orchards and covers an area of about 50,000 m<sup>2</sup>. The imagery was taken using a DJI FC6310R camera together with a DJI Phantom 4 RTK. The flight height was 100 m and the flight was conducted in July 2022 at about 10 a.m. under slightly cloudy skies. The raw imagery was processed using Agisoft Metashape in order to generate an orthomosaic, with a resulting GSD of approx. 2.8 cm. Border areas of the orthomosaic were cut out due to distortions, resulting in an orthomosaic of 8000×8000 pixels (around 224×224 m). A small section of the used dataset can be seen in Figure 4. For later processing, the orthomosaic is divided into 60 image tiles, each consisting of 1000×1000 pixels (approx. 28×28 m).



Fig. 4: Section of the test area near Erligheim, mainly consisting of cherry orchards

## 5 Experiments

Two crowd campaigns were performed, one for each work step. Users were only able to participate once per campaign.

### 5.1 First Crowd Campaign

Each of the 60 image tiles in our dataset was presented to 6 different crowdworkers, resulting in a total of 360 crowd acquisitions. Users were paid \$0.15 per 3 acquisitions, for a total cost of \$18 for the whole dataset, or \$0.30 per image tile.

### 5.2 Second crowd campaign

Processing the first crowd campaign led to a total of 125 individual trees that were detected. Smaller image sections, only containing these single trees, were created. These were presented to 15 different crowdworkers per tree, resulting in 1,875 total acquisitions at a cost of \$0.10 per 3 trees, for a total cost of \$62.5, or \$0.50 per tree.



## 6 Evaluation

### 6.1 First crowd campaign: Coarse tree positions

Ground truth and integrated values can be compared directly for all grid cells. No ground truth was available, therefore reference data were collected by experts within our institute.

All raster elements can be divided in the categories “true positive” (TP), “false positive” (FP), “true negative” (TN) and “false negative” (FN), see OLSEN & DELEN (1997). Using these categories, the values for precision, recall and F1 score can be calculated for a quantification of quality of the integrated classification. Table 1 shows these results, both before and after the integration.

Table 1: Mean precision, recall and F1 score of all image tiles before and after integration

	Before integration	After integration
Mean Precision	87,73%	89,01%
Mean Recall	66,86%	85,25%
Mean F1	79,11%	87,73%

The relatively good average precision of around 88% remains mostly unchanged through the integration. We do only see little improvement, but, more importantly, there is no decrease either. On the other hand, the rather low recall value of around 67% increases up to 85%, resulting in a significantly improved F1 score overall. Thus, we can conclude that a multi-fold acquisition combined with element-wise majority voting may only have a small effect on worker accuracy, but can significantly increase data completeness. Figure 5 shows the distribution of F1 scores before and after the integration and illustrates the impact of integration on the F1 score.

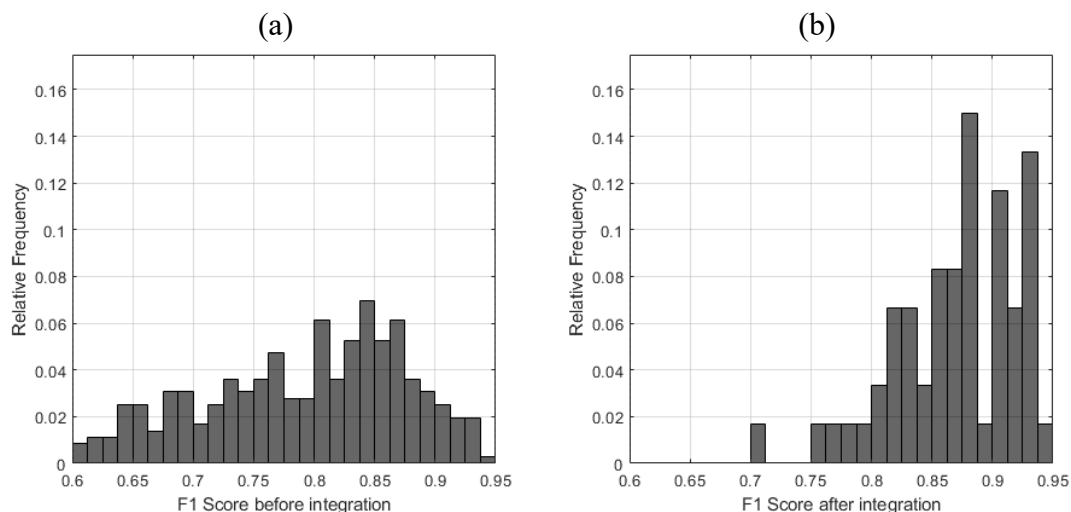


Fig. 5: Distribution of F1 scores. (a) Before integration and (b) after integration

### 6.2 Second crowd campaign: Precise tree outlines

For the evaluation of the polygon acquisitions, a ground truth is required as well. Since no independent reference data were available here either, we again had to rely on an expert-generated ground truth.

For each polygon acquisition, we calculated the Jaccard index to the respective ground truth. The Jaccard index calculates the similarity between two samples by calculating the Intersection over Union (JACCARD 1901), with 1 being a perfect similarity and 0 being no similarity at all. The mean calculated Jaccard index of all crowd acquisitions before any integration measures was 0.691, indicating an average similarity of 69.1% between acquisition and reference. This relatively low value can be attributed to the fact that no prior filtering of noisy data was performed, and crowd acquisitions in general tend to be noisy, visible in Figure 6b.

The integration approach described in section 5.2 is performed with all 125 trees of the dataset. An example can be seen in Figure 6, including all raw acquisitions (Figure 6b) and the integrated result (Figure 6c), using 8 as the binary vote threshold.

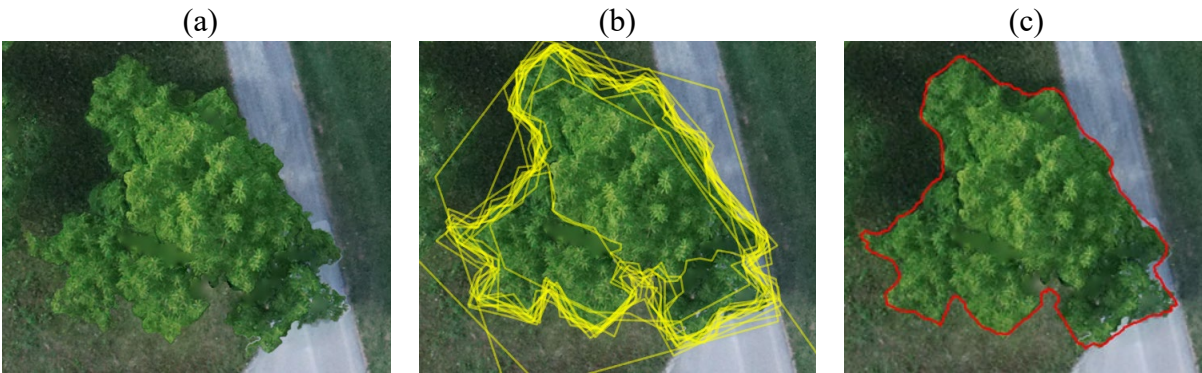


Fig. 6: Integration steps. (a) Raw image, (b) 15 Crowd acquisitions, (c) Integrated result

Figure 6 illustrates the effectiveness of our approach. While Figure 6b contains multiple noisy acquisitions, the integrated result consists of a clear outline, apparently free of noise, as can be seen in Figure 6c.

The Jaccard index can be calculated for all integrated polygons and then compared to the pre-integration values per tree, directly indicating the quality improvement achieved.

However, before integration, the binary vote threshold needs to be defined. Figure 3 in section 3.2 suggests that the integrated result is dependent on the threshold value. To show the influence of this binary vote threshold, we calculated all integrations for all 125 trees with all 15 possible values, their respective Jaccard index compared to the ground truth per tree, and the mean values. The result is shown in Figure 7.

As can be seen, best results are obtained with a threshold value of 7 with a resulting IoU of 0.834, compared to an IoU value of 0.691 without any integration measures. In this case, picking a value lower than 3 or higher than 11 will result in worse quality than not integrating the results at all. This can be explained in a fairly straightforward manner: Choosing the threshold value too low will include more acquisitions for the calculation of the integrated shape, even noisy ones, resulting in a decline in quality due to an increase of size of the integrated polygons: More grid cells outside the actual tree geometry are assigned to belong to the integrated polygon. Choosing the threshold too high on the other hand will exclude those grid cells not labeled by a vast majority of workers, leaving only a small pool of grid cells and resulting in a decline in quality due to an effect similar

to an erosion. The presumption that the quality of the integrated result is dependent on the picked binary vote threshold can therefore be confirmed.

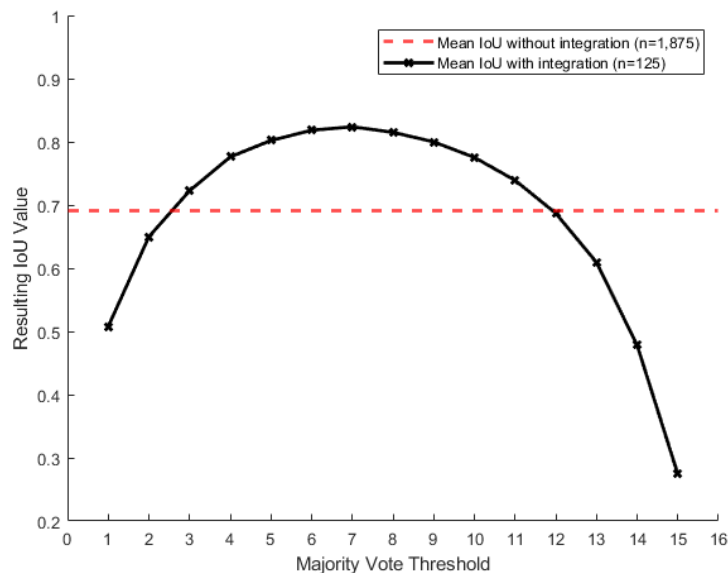


Fig. 7: Mean intersection over union (IoU) values of all trees for different binary vote thresholds

In our case, best IoU values are achieved with a vote threshold of 7, indicating the best improvement through integration. Not only does the mean IoU value increase by around 14 percentage points from 0.691 to 0.834 through the integration, the standard deviation of all observed values also reduces from a value of 0.198 before integration to a value of 0.119, equaling a reduction of around 40%. The distribution of the data values improves by a large margin, as can be seen in the histogram in Figure 8.

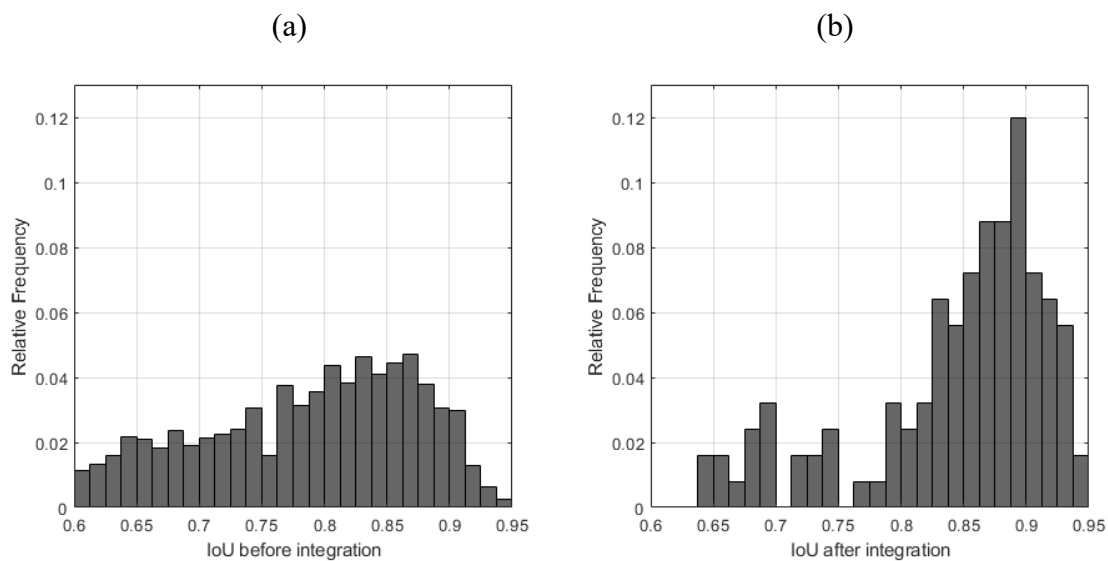


Fig. 8: Distribution of IoU values. (a) Before integration and (b) after integration

## 7 Conclusion

Using majority voting as integration measure, we were able to acquire a high-quality tree segmentation map with an average F1 score of nearly 88%, confirming that noisy data can be improved through repeated acquisitions with subsequent integration. Furthermore, we were able to prove this concept to be true even for irregular polygons when using a raster integration approach. The integrated polygons improve in quality, which can be confirmed not only visually, but also by Intersection over Union values.

The integrated polygons consist of a vast number of vertices and have complex geometries even when converted back to vector data. If smooth data are desired, line smoothing operations could be considered for cleaning up geometries. The high number of vertices could lead to either a more accurate geometry, since more details may be included, or a reduced accuracy due to higher noise. Further observations are required in order to investigate the influence of vertices number on the general geometry and to determine if there is a relationship with the binary vote threshold parameter.

## 8 Acknowledgement

Partially supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2120/1 – 390831618.

## 9 References

- ALLAHBAKHSI, M., BENATALLAH, B., IGJATOVIC, A., MOTAHARI-NEZHAD, H. R., BERTINO, E. & DUSTDAR, S., 2013: Quality control in crowdsourcing systems: Issues and directions. *IEEE Internet Computing*, **17**(2), 76-81.
- CHANDLER, J., PAOLACCI, G. & MUELLER, P., 2013: Risks and rewards of crowdsourcing marketplaces. *Handbook of human computation*, Springer Verlag, New York, 377-392.
- CHEUNG, J. H., BURNS, D. K., SINCLAIR, R. R. & SLITER, M., 2017: Amazon Mechanical Turk in organizational psychology: An evaluation and practical recommendations. *Journal of Business and Psychology*, **32**(4), 347-361.
- FINNERTY, A., KUCHERBAEV, P., TRANQUILLINI, S. & CONVERTINO, G., 2013: Keep it simple: Reward and task design in crowdsourcing. *Proceedings of the Biannual Conference of the Italian Chapter of SIGCHI*, 1-4.
- HADI H., YOWARGANA, P., ZULKARNAIN, M. T., MOHAMAD, F., GOIB, B.K., HULTERA, P., STURN, T., KARNER, M., et al., 2022: A national-scale land cover reference dataset from local crowdsourcing initiatives in Indonesia. *Scientific Data*, **9**(1), 1-15.
- HIRTH, M., HOFELD, T. & TRAN-GIA, P., 2011: Anatomy of a crowdsourcing platform-using the example of microworkers.com. *2011 Fifth international conference on innovative mobile and internet services in ubiquitous computing*, 322-329.
- HOSSAIN, M., 2012: Users' motivation to participate in online crowdsourcing platforms. *2012 International Conference on Innovation Management and Technology Research*, 310-315.

- JACCARD, P., 1901: Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bull Soc Vaudoise Sci Nat*, **37**, 547-579.
- OLSON, D. L. & DELEN, D., 2008: *Advanced Data Mining Techniques*, Springer Verlag, 1st edition, 138.
- PILZ, D. & GEWALD, H., 2013: "Does money matter? Motivational factors for participation in paid- and non-profit-crowdsourcing communities.". *Proceedings of the 11th International Conference on Wirtschaftsinformatik (WI)*, **1**, 577-592.
- PUTTINAOVARAT, S., SAELIW, A., PRUITIKANEE, S., KONGCHAROEN, J., CHAI-ARAYALERT, S. & KHAIMOOK, K., 2022: Flood Damage Assessment Geospatial Application Using Geoinformatics and Deep Learning Classification. *International Journal of Interactive Mobile Technologies*, **16**(21).
- RYAN, R. M. & DECI, E. L., 2000: Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology*, **25**(1), 54-67.
- SARALIOGLU, E. & GUNGOR, O., 2020: Crowdsourcing in remote sensing: A review of applications and future directions. *IEEE Geoscience and Remote Sensing Magazine*, **8**(4), 89-110.
- SARALIOGLU, E. & GUNGOR, O., 2022: Crowdsourcing-based application to solve the problem of insufficient training data in deep learning-based classification of satellite images. *Geocarto International*, **37**(18), 5433-5452.
- STONEBRAKER, M. & REZIG, E., 2019: "Machine learning and big data: What is important?". *IEEE Data Eng. Bull.*, **42.4**, 3-7.
- WALTER, V., KÖLLE, M. & YIN, Y., 2020: Evaluation and optimisation of crowd-based collection of trees from 3D point clouds. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, **4**, 49-56.
- WALTER, V., KÖLLE, M., COLLMAR, D. & ZHANG, Y., 2021: A two-level approach for the crowd-based collection of vehicles from 3D point clouds. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, **4**, 97-104.
- WHANG, S. E. & LEE, J. G., 2020: Data collection and quality challenges for deep learning. *Proceedings of the VLDB Endowment*, **13**(12), 3429-3432.
- YUEN, M. C., KING, I. & LEUNG, K. S., 2012: Task recommendation in crowdsourcing systems. *Proceedings of the first international workshop on crowdsourcing and data mining*, 22-26.
- ZHANG, J., WU, X. & SHENG, V. S., 2016: Learning from crowdsourced labeled data: a survey. *Artificial Intelligence Review*, **46**(4), 543-576.

# Satellite Network Bavaria – Mission and Data Processing

MICHAEL GREZA<sup>1</sup>, LUDWIG HOEGNER<sup>2</sup>, PHILIPP-ROMAN HIRT<sup>1</sup>,  
ROBERT ROSCHLAUB<sup>3</sup> & UWE STILLA<sup>1</sup>

*Abstract: Agriculture and silviculture as well as governmental institutions are facing major challenges concerning climate change and species conservation. Intensifying extreme weather events add pressure to establish a sustainable management which requires optical geodata with a high temporal and spatial resolution. This also encompasses new demands in the private sector.*

*Existing missions prove that small satellites can fulfill this demand in an economical manner. Especially the CubeSat standard enables the development of satellites outside of big space agencies and therefore specialized missions that take regional demands into account.*

*To fulfill its sovereign tasks in the aforementioned disciplines with their current and future challenges, the Bavarian state initiated studies to determine the potential of a small satellite mission. The resulting demonstrator mission comprises a formation of five 6U CubeSats, each equipped with a multispectral camera, to monitor state territory with high resolution and frequency. The mission is executed as a research alliance between the Bavarian surveying administration, the Technical University of Munich, and the Center for Telematics Würzburg. It encompasses the entire process chain, from mission design and satellite manufacturing up to data processing. The publication describes the mission concept, focusing on data processing strategies. Further, a simulator for artificial satellite imagery is presented.*

## 1 Mission Introduction

Earth observation imagery demand and availability is growing rapidly, as pressures from climate change and global market dynamics rise, as well as its list of valuable applications. Conventional earth observation satellites, e. g. Sentinel-1 to -6, are a potent data source but have downsides in long development cycles at a high cost and come with a tradeoff between ground sampling distance (GSD) and ground coverage. Only highly funded agencies can launch singular exemplars. In contrast, with low cost and fast development cycles, CubeSats are an alternative for stakeholders that enables them to tailor satellite missions to their unique needs while remaining affordable. Often, their drawbacks because of the small form factor can be mitigated by adapting mission parameters and launching larger constellations, still costing less than a singular conventional earth observation satellite.

As the most prominent contender, Planet Labs' daily repetition cycle for an almost global, multispectral coverage through the PlanetScope constellation (PLANET TEAM 2017) satisfies the demand for readily available, high-resolution data. Still, its mission design has drawbacks, e.g.,

---

<sup>1</sup> Technische Universität München, Photogrammetrie und Fernerkundung, Arcisstr. 21, 80333 München, E-Mail: [michael.greza, philipp.hirt, stilla]@tum.de

<sup>2</sup> Hochschule München, Fakultät 08 Labor für Photogrammetrie und Fernerkundung, Karlstr. 6, 80333 München, E-Mail: ludwig.hoegner@hm.edu

<sup>3</sup> Landesamt für Digitalisierung, Breitband und Vermessung, Alexandrastr. 4, 80538 München, E-Mail: robert.roschlaub@ldbv.bayern.de

concerning image acquisition consistency and data sovereignty. Creating similar acquisition parameters like sun incident angle i.e., constant local revisit time, or utilizing the same satellite for a region repeatedly to enhance change detection algorithms radiometrically, are desirable in remote sensing. In other terms, indelible accessibility to this security-relevant and now vital data is not guaranteed with foreign private companies.

The Bavarian surveying administration (BVV) conducts biannual aerial image acquisition campaigns (“Bayernbefliegung”) to obtain very-high-resolution imagery of the free state. The 20 cm GSD RGBIPAN products can be utilized for sophisticated use cases but lack a high temporal resolution, ruling out short- and medium-term change monitoring. Exemplary applications that are not feasible are intraseasonal agricultural and silvicultural products like growth monitoring or natural disaster damage assessment. Increasing flight frequency to a weekly timeframe would result in exploding costs and is therefore not viable.

To assess the potential of a state-owned CubeSat solution for this gap, the BVV conducted preliminary studies together with TUM Photogrammetry and Remote Sensing (PF), Astronomical and Physical Geodesy (APG) and Center for Telematics (“Zentrum für Telematik”, ZfT) Würzburg. State agencies were involved to identify stakeholder requirements. Core needs were state-wide coverage, more than five years of lifetime, a repeat cycle of less than five days with consistent local revisit times, a GSD below five meters and a spectral range that supports agricultural and silvicultural applications. Mission parameters and design were derived with the help of simulators.

As state-wide coverage requires a larger satellite constellation, a precursory demonstrator mission was commissioned to verify mission, orbit and satellite design as well as the data processing chain. This demonstrator mission is outlined below, starting with important mission design elements, followed by aspects of the processing chain and composition of the PF simulator. Parameters still might change in detail.

## 2 Preliminary Mission Design

### 2.1 Satellites

As this publication focuses on the description of the simulation and data processing systems, satellite engineering and design will only be described in an overview. The constellation encompasses five 6U CubeSats identical in construction. Each carries a multispectral camera outlined in chapter 2.2 as its main payload. To be able to achieve the desired lifetime of at least five years and to maintain the orbit configuration presented in chapter 2.3, they are equipped with a chemical engine. X-band, S-band as well as UHF antennas are used for data downlink, uplink, and TT&C. GNSS antennas provide satellite location data and solar panels on all unoccupied surfaces provide power.

This mission has high demands concerning the ADCS’s capabilities for a CubeSat. To keep the total number of satellites needed for a gap-free coverage of Bavaria as low as possible and ensuring an overlap while being economically efficient, satellite orientation has to be determined with high precision. As a coarse measuring instrument, a sun sensor is used. For high precision attitude

determination two perpendicularly mounted star trackers are installed. Attitude is controlled via magnetorquers as well as reaction wheels.

One of the biggest challenges in this mission's satellite design is to fit all needed subsystems into the frame as build and launch costs grow with constellation size. A 3U or 4U chassis does not provide the necessary space for all subsystems. This constrain is also reflected in the orbit design, as the available space for the camera's optics limits GSD which can then only be improved by lower orbit heights and therefore more narrow swath widths.

Final satellite assembly is taking place at ZfT.

## **2.2 Payload**

The main payload is an eight-channel multispectral pushbroom CMOS line scanner. Its spectral filters are selected to accommodate for mainly agricultural and silvicultural application purposes in a range of 450 – 900 nm as well as a PAN channel. Each spectral band can use up to 32 TDI stages for improved illumination and image quality in the 10-bit pixels. The lens system offers a field of view of  $2.22^\circ$  at a focal length of 580 mm.

## **2.3 Orbit**

The constellation of five satellites will be flown in a string of pearls configuration on a sun synchronous low earth orbit (LEO) with a repeat rate of roughly three days, passing Bavaria before noon for optimal atmospheric and constant lighting conditions. Once within each repeat cycle, the satellites consecutively capture overlapping image stripes of Bavaria. The center satellite will pass over the east of Munich on a near-polar orbit at a mean height of 460 km, resulting in a mean swath width of roughly 18 km at 4 m GSD. After the satellites acquire the images in imaging mode, they are turned to face the ground stations for data transfer. When data transmission is complete, they are turned back to their imaging configuration. All orbit and orientation maneuvers are executed after imaging, so none are executed during the image capturing overpasses, ensuring a stable and vibration-free environment.

Orbits are designed and calculated by APG and ZfT.

# **3 Preliminary Data Processing**

## **3.1 Data Processing Chain**

The amount of data that is generated by the satellites within a repeat frequency of roughly three days does not allow for semiautomated processing, whether in the demonstrator or in the full extension of the mission. Currently, the BVV processing chain for the Bayernbefliegung contains manual steps to interactively select key points. A fully automated processing chain is being built to ensure timely data handling within the satellites' revisit time.

At data processing level 0, the consistency-checked raw image data is transferred from the ground stations. The IT service center Bavaria is providing the infrastructure for storage and data handling. TUM, BVV and ZfT all provide algorithms for the successive data processing chain. The image preprocessing system for level 1 is laid out in chapter 3.2. Details on processing done by ZfT and APG or subsequential products are not further described within the scope of this publication.



### 3.2 Data Processing Level 1 A&B

After the initial preprocessing of the bit-wise raw data, geometric and radiometric corrections are applied. Radiometric calibration data is provided by the camera manufacturer. Additionally, PF and ZfT acquire ground truth surface correction data collected at agricultural and silvicultural research sites of TUM.

For all processing elements a backup solution is designed to ensure the processability of the real satellite data under differing, adverse conditions. The alignment, coregistration and georeferencing steps are done iteratively to refine the results.

As the camera is a pushbroom system, correct alignment of the individual image rows is essential. To achieve this, both a deep neural network and a SIFT feature detector (LOWE 2004) are used to identify tiepoints in overlapping image strips with high confidence on a sub-pixel level. The tie points and a DTM together with the attitude and orbit information provided by the satellites are used within a bundle block adjustment system to generate a first corrected, more precise set of attitude and orbit information to realign the individual image rows for corrected strips.

The image strips are then divided into blocks less than 75 km in length for further processing at the BVV. This is to ensure that earth's curvature does not impose restrictions on the 2D intermediate processing. Building centers are extracted to find a mapping to cadastral data for a first georeferencing of the image blocks. A neural network for building center extraction on aerial imagery (ROSCHLAUB et al. 2020) is currently being adapted to simulation data for our mission. The extracted centroids are fused into settlement clusters in resolutions of 40, 400 and 4.000 m. Then, a transformation from the extracted cluster centroids to the settlement cluster centroids of the cadaster is calculated (ROSCHLAUB et al. 2023) to obtain preliminary transformation parameters. After this rough transformation, the alignment can be further refined at the full 4 m resolution as coarse misalignment errors with individual buildings are now precluded.

A backup system for the coregistration of key points uses density maps of extracted and cadastral building centroids and calculates transformation parameters via cross correlation. This is a simple and stable system but due to its computational cost not preferred.

As an alternative to the above-mentioned process, a neural network that directly coregisters the individual image rows onto an already georeferenced orthoimage will be tested. This will also utilize a DTM within the coregistering chain.

## 4 Mission Simulator

### 4.1 Overview

To be able to determine the mission requirements more precisely and to verify assumptions on the performance of the overall system concerning data processing possibilities, a data simulator is being built. It is designed to assess the impact of design and part choices on the expected resulting data throughout the processing chain and generate test data. APG, ZfT and BVV also implemented testbeds for their respective focus areas. The current elements of its PF part are discussed below.

## 4.2 Virtual Satellite Camera

A core element of the simulation is the ability to produce artificial satellite data which resembles the expected real mission results and therefore shows the impact of design choices and parameters on the produced raw data and subsequent processing products. As the virtual satellite camera's detailed underlying methodology is part of a separate DGPF publication (LENZ & GREZA 2023), the following section will focus on its role within the whole simulator.

High resolution TrueDOPs act as a base image from which the artificial satellite images are created. The BVV provides PF with TrueDOP40 imagery from its Bayernbefliegung campaign for this purpose. For its core functionality, a virtual camera flies over the base imagery and captures it with respect to sensor specs, camera parameters, the orientation of the satellite, field of view and radiometric behavior. Its path is determined by propagated orbits that were created by APG. During the capturing process, the base image pixels are projected onto the sensor pixels and degraded through filtering and modulation transfer functions. This produces more realistic image lines than simple downsampling from the 0.4 m base GSD to the 4 m GSD of the satellite. The next iteration of the virtual satellite camera will also make use of a high resolution DTM to further improve the imaging geometry, which is especially relevant in mountainous regions as lines of sight are heavily impacted in these areas.

As there are no ground truth images available yet, the tuning of radiometric and geometric postprocessing is currently bound to qualitative assessment of the images by comparing to images of different satellite systems, i. e., Sentinel-2 (SENTINEL HUB 2023), Planet Dove (PLANET TEAM 2017), and Landsat 8 & 9 (USGS 2023). To tackle this, several radiometric and geometric calibration tests are tasked with the engineering model camera and its lens system. Their results will be used as more accurate parameters and enable a quantitative radiometric assessment of the virtual camera.

The simulated image strips are then used to test the processing chain that is laid out in chapter 3. Figure 1 shows an exemplary cutout of a virtual overpass. The most important tests for the mission design are the production of strips with differing overlap values to determine the minimum desired overlap and the introduction of tumbling, as well as attitude and position inaccuracies. These verify the constellation design and satellite subsystem requirements. In figure 1, a test with mismatching ground speed and image acquisition frequency results in a slight along-track distortion.



Fig. 1: Simulated satellite image with mismatching line rate at 4 m GSD derived from DOP40

### 4.3 Image Degradation

#### 4.3.1 Optics and Sensor Degradation

An automatic processing chain must be robust against adverse conditions. To find breaking points within the system, the simulator imagery can also be further altered to reflect potential threats to the processability.

One option, which was implemented with support from Ramesse Zatti, is to add sensor noise and pixel faults that could emerge in orbit as well as simulate the effects of particles on the lens. One of the driving causes for degradation in LEO is atomic oxygen, which due to its reactive nature can erode spacecraft surfaces and create particles that settle on the sensor and lenses. Several publications on atomic oxygen damage in LEO provide modelled and measured values for various surface orientations in relation to flight direction. As it was not possible to generate a consistent model from their results, three distinctive values were chosen for our simulations. Useful values from the Long Duration Exposure Facility and the Tropical Rainfall Measuring Mission (TRMM) were taken as reference. TRMM operated during a solar maximum, resulting in high values of atomic oxygen. A total atomic oxygen fluence of  $8.9 \cdot 10^{22}$  atoms per  $\text{cm}^2$  was calculated (CHEN 2001).

Value 1: With a front-facing camera, we assume that 95 % of atomic oxygen fluence results in damages. (PETERS et al. 1986)

Value 2: With a side-facing camera, we assume that 4 % of atomic oxygen fluence results in damages (BANKS et al. 2004).

Value 3: With a rear-facing camera, we assume an atomic oxygen fluence of  $10^4$  (CHEN 2001).

For first tests we assumed a pixel failure rate of 0.5 % per year like the TRMM experienced in LEO based on (CHEN 2001) and adapted this to the three standard camera orientations. It is to note that these 0.5 % is a combination of atomic oxygen effects as well as charged particle interference.

As expected, test results show that a rear-facing camera during the satellites' maneuver configuration is to prefer over a front-facing orientation.

Further particle sources apart from atomic oxygen are material outgassing, venting, leaks, and propellant plumes. Together with residuals from assembly and launch, contaminations and dust depositions occur on the optics, resulting in smaller transmissive capabilities of the lens system. This effect is applied to the image in figure 2. These can become a relevant element over the planned mission lifetime over five years. The impact of the string of pearls constellation configuration executing maneuvers resulting in propellant plumes have to be analyzed and modelled further to derive recommendations for camera orientation during the maneuvers. It is assumed that a trailing camera will also be favourable.



Fig. 2: Simulated satellite image including heavy degradation for system failure tests

#### 4.3.2 Cloud Generation

A second capability of the image degradation simulator is the possibility to generate artificial clouds within the images. As parts of the image processing chain rely on tie points and ground control points derived from deep neural networks, low visibility due to high cloud coverage can result in failure of the high precision processing. To be able to find these breaking points and create more resilient models that can cope with inter alia lighter fog or partly cloudy areas, a cloud generator was implemented with support from Yifan Xu. By using Perlin noise (PERLIN 2002), fog as well as cirrus, stratus (figure 3) and cumulus clouds can be generated with differing coverage and wind orientations.



Fig. 3: Exemplary simulated satellite image containing stratus clouds

## 5 Outlook & Discussion

### 5.1 Future Data Products

As the design phase is ending, the focus of the teams shifts from satellite and mission design towards assembly, component testing and the implementation of data processing on different levels. For the latter, the project team is working with Bavarian administrative bodies to develop new application systems for the data processing levels 2, 3 and 4.

More common target applications in agriculture and silviculture are precision farming techniques like yield prediction, plant health monitoring and disease detection as well as subsidy monitoring and damage assessment.

Further, rapidly available reconnaissance data after natural disasters are of high value and are a planned product outside of the routine applications. For this, a tasked spotlight mode can be tested within the demonstrator mission. This enables an even higher acquisition frequency and the generation of DEMs, for example after landslides.

### 5.2 Mission outlook & Discussion

As this is a demonstrator mission, the insights gained throughout development and especially during satellite operations will set the base for a follow-up full extension mission which will cover the whole state of Bavaria with its satellites. Especially the need for a high precision AOCS creates engineering challenges for a CubeSat system. Without real data available, the demonstrator mission is the only way to verify design choices for a full extension. AOCS precision directly correlates with the number of satellites needed for the full coverage of Bavaria and therefore cost. CubeSat constellations have impressive capabilities at a low cost, but some caveats remain. Design restrictions because of their limited space can be easily seen in the size limits of feasible optics and in data rates. Even with its high-bandwidth antennas, data transfer imposes restrictions on possible image strip lengths. With its full extension, the mission will cover the whole of Bavaria but transferring significantly longer strips (e. g. also covering parts of Hesse, Thuringia, Northrhine-

Westphalia, and Lower Saxony) is not possible within the revisit time in current satellite specifications. Expanding the mission and satellite design after a successful demonstration should be discussed to further expand the capabilities of the system.

## 6 Bibliography

- BANKS, B., MILLER, S. & DE GROH, K., 2004: Low earth orbital atomic oxygen interactions with materials. 2nd International Energy Conversion Engineering Conference, 2004, 5638.
- CHEN, P., 2023: Contamination Effects due to Space Environmental Interactions. AIAA 2005-0095. 39<sup>th</sup> Aerospace Sciences Meeting & Exhibit, 08.-11. January 2001.
- LENZ, N. & GREZA, M., 2023: Simulation of Earth Observation Data Utilizing a Virtual Satellite Camera. Publikationen der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation e.V., 43. Wissenschaftlich-Technische Jahrestagung der DGPF, 22.-23. März 2023 in Munich, **31**, 50-60.
- LOWE, D., 2004: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, **60**, 91-110.
- PERLIN, K., 2002: Improving Noise. Proceedings of the 29th annual conference on Computer graphics and interactive techniques, 681-682.
- PETERS, P.N., GREGORY, J.C. & SWANN, J.T., 1986: Effects on optical systems from interactions with oxygen atoms in low earth orbits. Applied Optics, **25**(8), 1290-1298.
- PLANET TEAM, 2017: Planet Application Program Interface: In Space for Life on Earth. <https://api.planet.com>, last access 30.01.2023.
- ROSCHLAUB, R., LI, Q., AUER, S., MÖST, K., GLOCK, C., SCHMITT, M., SHI, Y. & ZHU, X., 2020: KI-basierte Detektion von Gebäuden mittels Deep Learning und amtlichen Geodaten zur Baufallerkundung, **145**(3), 180-189.
- ROSCHLAUB, R., GLOCK, C., ZERNDL, M., MAIER, A., MÖST, K., HIRT, P.-R. & GREZA, M., 2023: Grundlagen zur automatisierten Ermittlung hochgenauer Passpunkte für CubeSat-Satellitenbilder mittels Deep-Learning-gestützter Delaunay-Triangulation basierend auf Gebäudedaten des Liegenschaftskatasters. ZFV - Zeitschrift für Geodäsie, Geoinformation und Landmanagement, **148**(2).
- SENTINEL HUB, 2023: Modified Copernicus Sentinel-2 data, 2017-2023. <https://www.sentinel-hub.com/>, last access 23.01.2023.
- USGS, 2023: Landsat 8 & 9 imagery courtesy of the U.S. Geological Survey, 2021-2023. <https://earthexplorer.usgs.gov/>, last access 12.01.2023.

# Modelling of GNSS Positioning Errors in a GNSS/INS/LiDAR-integrated Georeferencing

FLORIAN PÖPPL<sup>1</sup>, MARTIN PFENNIGBAUER<sup>2</sup>, ANDREAS ULLRICH<sup>2</sup>,  
GOTTFRIED MANDLBURGER<sup>1</sup>, HANS NEUNER<sup>1</sup> & NORBERT PFEIFER<sup>1</sup>

*Abstract: Kinematic laser scanning is an efficient and highly accurate method for the acquisition of 3D topographic data. A primary task in kinematic laser scanning is the transformation of the laser scanner measurements from a local scanner coordinate system to a global georeferenced coordinate system. This requires knowledge of the scanners' trajectory (position and orientation over time). The trajectory is typically computed via Kalman filtering of Global Navigation Satellite System (GNSS) and Inertial Navigation System (INS) measurement data. However, this trajectory often exhibits significant errors, which become apparent when point clouds acquired at different times overlap spatially. In these overlapping areas, corresponding points may be exploited in a subsequent strip adjustment to improve the trajectory and the system calibration and therefore the final point cloud. However, ignoring the raw GNSS and INS measurements and their statistical properties can lead to over-optimistic estimates and deformation of the laser point cloud. In this contribution, a method is presented which integrates the GNSS position, raw inertial measurements and laser scanner correspondences into one adjustment and explicitly models the time-correlated stochastic errors of the GNSS positioning solution. This method for direct georeferencing considerably reduces the discrepancies between overlapping point clouds, as demonstrated on an airborne laser scanning dataset in comparison with an existing state-of-the-art strip adjustment implementation.*

## 1 Georeferencing for Kinematic Laser Scanning

Georeferencing is an important task in kinematic laser scanning, since most use-cases require the acquired point cloud to be in a suitable georeferenced coordinate system. The default way of achieving this involves GNSS (Global Navigation Satellite System), which provides absolute positioning, and INS (Inertial Navigation System), which provides high frequency relative position and orientation. Classically, GNSS and INS are fused via Kalman filtering to obtain the platform trajectory. When the geometric relation between platform and laser scanner is known, the laser measurements can be transformed from the scanner's coordinate system into the platforms' coordinate system and then into the earth-referenced coordinate system of the trajectory (Fig. 1). These errors in turn propagate through the georeferencing and result in errors in the point clouds, which become visible as discrepancies in areas that are scanned several times. This can be partially rectified by identifying correspondences in overlapping areas and introducing them as observations

---

<sup>1</sup> Technische Universität Wien, Department für Geodäsie und Geoinformation,  
Wiedner Hauptstraße 8, 1040 Wien, Österreich,  
E-Mail: [florian.poeppel, gottfried.mandlbuerger, hans.neuner, norbert.pfeifer]@geo.tuwien.ac.at

<sup>2</sup> RIEGL Laser Measurement Systems GmbH, Riedenburgstrasse 48, 3580 Horn, Austria,  
E-Mail: [mpfennigbauer, aullrich]@riegl.com

in a strip adjustment (Fig. 2, FILIN & VOSSELMAN 2004; KAGER 2004), thereby correcting errors in position and/or orientation and possibly the laser scanner calibration.

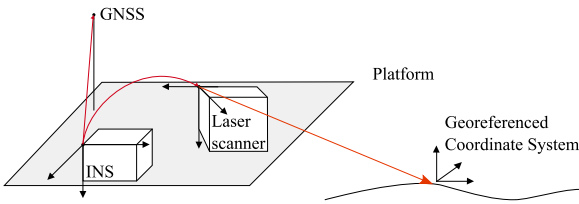


Fig. 1: GNSS antenna, INS and laser scanner mounted rigidly on the multi-sensor platform

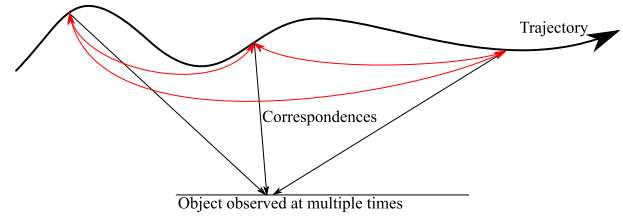


Fig. 2: Corresponding points on the ground form constraints on the trajectory at the acquisition times

However, such an approach is typically based on correcting the existing trajectory, and does not take into account the underlying GNSS or INS measurements or their stochastic properties. To allow correction of time-varying errors, a flexible trajectory correction model (e.g. splines) must be used, leading to an increase in the number of parameters. Without sufficient ground control this results in overparametrization and overfitting of the model as well as in systematic deformation of the point cloud (GLIRA et al. 2016).

In this work, a different approach is presented: The position observations from an initial GNSS-only processing are combined with the inertial measurements in a non-linear least-squares adjustment to produce an initial trajectory, which is used for a first georeferencing of the laser data. Then, correspondences are extracted from overlaps in the laser point cloud and used as input to a single least-squares adjustment together with the GNSS positions and inertial measurements. This allows simultaneous determination of time-constant calibration parameters, time-varying errors in the inertial sensors, and the trajectory (Tab. 1). Conceptually, this is a tight coupling of laser scanner measurements with the INS and a loose coupling with the GNSS, where information from the laser measurements helps to stabilize the INS drift. However, loose coupling of the GNSS poses its own challenges. The errors in the GNSS positioning may be correlated in time and in space due to a) the processing itself – usually via Kalman filter – and b) not fully modelled systematic errors such as ionospheric or tropospheric effects. Depending on the strength of the correlations, disregarding the time-correlated nature of the GNSS position observations leads to overly optimistic or biased parameter estimates. We propose explicitly accounting for the time-correlation in the stochastic model by assuming an additional stochastic error term modelled by a first-order Gauss-Markov process. This is similar to how inertial sensor errors are commonly modelled (FARRELL et al. 2022). Applying this to GNSS errors is not new in the context of Kalman filters (NIU et al. 2018), but has to our knowledge never been done in the context of batch least-squares estimation in general or strip adjustment in particular. We also present a simple yet effective way of estimating the relevant parameters of the stochastic model.

In Section 2, we will first discuss the methodology and then in Section 3 demonstrate our approach by applying it to an airborne laser scanning dataset. This dataset was acquired with a RIEGL VUX-160<sup>23</sup> laser scanner. The scanner features three differently oriented mirror facets, resulting in three *views*: nadir, +10 degrees forward, and -10 degrees backward (Fig. 3). A comparison of the strip



differences of a) strip adjustment with constant per-strip trajectory correction, b) our method with uncorrelated GNSS error model and c) our method with time-correlated GNSS error model shows significantly reduced discrepancies for our proposed approach (c).



Fig. 3: RIEGL VUX-16023 scan pattern with nadir/forward/backward views

## 2 Integrated Georeferencing with Time-Correlated GNSS Errors

### 2.1 GNSS/INS/LiDAR-Integrated Georeferencing

In this section, we describe the functional and stochastic models for the trajectory estimation using pre-processed GNSS positions, raw inertial measurements from the INS and correspondences extracted from the LiDAR point cloud. This approach employs batch non-linear least-squares estimation, where the discrepancies between the model predictions and a given set of measurements are minimized. The least-squares minimization is solved by non-linear optimization techniques, namely the Levenberg-Marquardt algorithm as implemented in the Ceres library (AGARWAL et al. 2022).

Figure 4 shows the processing pipeline for the GNSS/INS/LiDAR integration. The main steps are:

1. Computing an initial trajectory from GNSS/INS data via batch least-squares adjustment, using a Kalman filter to initialize trajectory parameters for the non-linear optimization.
2. Georeferencing of the LiDAR data with the initial trajectory and calibration parameters to obtain an initial point cloud for identifying correspondences.
3. Computing an improved trajectory and calibration with a batch least-squares adjustment of all available data: GNSS, INS and LiDAR.
4. Georeferencing of the LiDAR data with the improved trajectory and calibration to obtain the final point cloud.

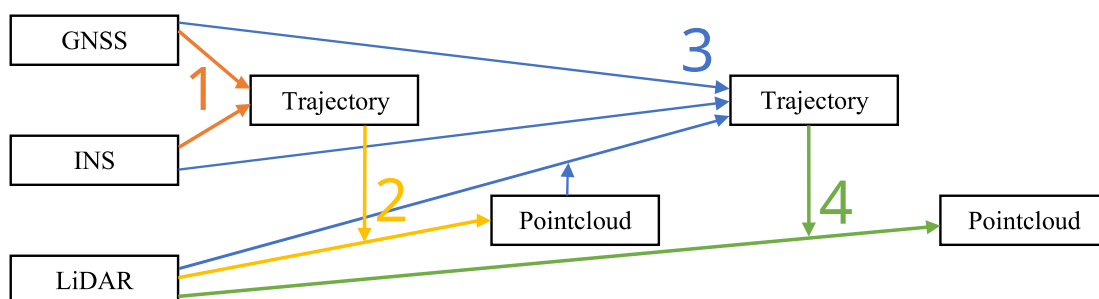


Fig. 4: Processing pipeline for GNSS/INS/LiDAR integration

The rest of this section describes the details of the least-squares adjustment in step 1 and step 3. A summary of the parameters occurring in the least-squares estimation is given in Tab. 1.

Tab. 1: Parameters for the least-squares estimation.

Trajectory	<ul style="list-style-type: none"> <li>• Position: cubic spline with coefficients <math>\mathbf{x}_p \in \mathbb{R}^{3x(n+2)}</math>, provides a continuous position <math>\mathbf{p}(t, \mathbf{x}_p) \in \mathbb{R}^3</math> of the platform in the local coordinate system.</li> <li>• Orientation: quadratic quaternion spline with coefficients <math>\mathbf{x}_o \in \mathbb{R}^{4x(n+1)}</math>, provides a continuous platform orientation <math>\mathbf{R}(t, \mathbf{x}_o) \in \text{SO}(3)</math>, i.e., the rotation from the platform coordinate system to the local coordinate system.</li> </ul> <p>Position and orientation are with respect to a <i>local tangent plane coordinate system</i>. Spline nodes are set at the sample times of the <math>n</math> inertial measurements. To simplify presentation, we define the platform coordinate system to be identical to the INS coordinate system.</p>
GNSS	<ul style="list-style-type: none"> <li>• Antenna lever-arm <math>\mathbf{l}_a \in \mathbb{R}^3</math>: antenna position in platform coordinate system.</li> <li>• GNSS bias <math>\mathbf{b}_g \in \mathbb{R}^{3xm}</math>: time-varying bias in xyz coordinates of the local coordinate system, modelled as a Gauss-Markov process. Bias parameters are estimated for all <math>m</math> GNSS measurement.</li> </ul>
INS	<ul style="list-style-type: none"> <li>• Accelerometer bias <math>\mathbf{b}_f \in \mathbb{R}^{3xm}</math>: time-varying random walk process.</li> <li>• Accelerometer scale-factor <math>\mathbf{s}_f \in \mathbb{R}^3</math>: time-constant.</li> <li>• Gyroscope bias <math>\mathbf{b}_\omega \in \mathbb{R}^{3xm}</math>: time-varying random walk process.</li> <li>• Gyroscope scale-factor <math>\mathbf{s}_\omega \in \mathbb{R}^3</math>: time-constant.</li> </ul> <p>Time-varying INS biases are estimated at the sample times of the <math>m</math> GNSS measurement and linearly interpolated in-between.</p>
LiDAR	<ul style="list-style-type: none"> <li>• Scanner boresight angles <math>\boldsymbol{\kappa} \in [0, 2\pi)^3</math>: defines the rotation <math>\mathbf{R}(\boldsymbol{\kappa})</math> from the scanner coordinate system to the platform coordinate system.</li> <li>• Scanner lever-arm <math>\mathbf{l}_s \in \mathbb{R}^3</math>: scanner origin in the platform coordinate system.</li> </ul>
Objects	<ul style="list-style-type: none"> <li>• Plane offset and slope parameters <math>(o, z_1, z_2) \in \mathbb{R}^3</math> per object plane.</li> </ul>

**Trajectory model:** The trajectory is modelled as a continuous-time function. Standard cubic B-Splines are used for position and quadratic quaternion splines (KIM et al. 1995) for orientation. The continuous-time representation allows to evaluate the trajectory at any given time  $t$ . This is especially relevant for the high-frequency LiDAR data, where estimating the position and orientation for each emitted laser shot is unfeasible. In contrast to the spline-correction approach used in (GLIRA et al. 2016), the splines here are used to model the trajectory itself, with spline parameters introduced at the frequency of the inertial measurements. The aforementioned problem of overparametrization and block deformation is addressed by incorporating the raw inertial measurements and the GNSS position measurements in the strip adjustment, allowing for a highly flexible trajectory model without reducing redundancy.

**GNSS measurement model:** The GNSS is integrated in a loosely coupled manner. The raw measurements (pseudo-range, carrier-phase, Doppler) are first processed using a carrier-phase differential kinematic GNSS solution, in this case using the software package RTKLib (TAKASU & YASUDA 2009) in post-processing mode. The GNSS processing provides timestamped positions  $\tilde{\mathbf{p}}$  and estimates of their precision. The predicted position of the GNSS antenna at time  $t$  can be derived directly from platform position  $\mathbf{p}(t, \mathbf{x}_p)$ , orientation  $\mathbf{R}(t, \mathbf{x}_o)$  and GNSS lever arm  $\mathbf{l}_a$ . The measurement is assumed to be contaminated with an additive error  $\epsilon_p$ . In total, the GNSS measurement equation for a measurement at time  $t$  is given by

$$\underset{\text{measurement}}{\tilde{\mathbf{p}}} = \underbrace{\mathbf{p}(t, \mathbf{x}_p) + \mathbf{R}(t, \mathbf{x}_o)\mathbf{l}_a}_{\text{model}} + \underset{\text{error}}{\epsilon_p}. \quad (1)$$

**Inertial measurement model:** The measurements  $\tilde{\mathbf{f}}, \tilde{\boldsymbol{\omega}}$  of the specific force  $\mathbf{f}$  and angular rate  $\boldsymbol{\omega}$  at time  $t$  include systematic errors (bias  $\mathbf{b}_f/\mathbf{b}_\omega$  and scale factor  $\mathbf{S}_f/\mathbf{S}_\omega$ ) and additive noise  $\epsilon_f/\epsilon_\omega$

$$\begin{aligned} \tilde{\mathbf{f}} &= (\mathbf{I}_3 + \mathbf{S}_f)\mathbf{f} + \mathbf{b}_f + \epsilon_f \\ \tilde{\boldsymbol{\omega}} &= \underbrace{(\mathbf{I}_3 + \mathbf{S}_\omega)\boldsymbol{\omega} + \mathbf{b}_\omega}_{\text{model}} + \underset{\text{error}}{\epsilon_\omega}. \end{aligned} \quad (2)$$

Non-orthogonality is not considered, thus  $\mathbf{S}_f = \text{diag}(\mathbf{s}_f)$  and  $\mathbf{S}_\omega = \text{diag}(\mathbf{s}_\omega)$ . The predicted specific force  $\mathbf{f}$  and angular velocity  $\boldsymbol{\omega}$  are computed via inertial navigation equations (GROVES 2013; PÖPPL et al. 2023) and depend on platform position  $\mathbf{p}(t, \mathbf{x}_p)$  and orientation  $\mathbf{R}(t, \mathbf{x}_o)$ .

**LiDAR measurement model:** The LiDAR-derived measurements are based on detecting and matching planar features in the point clouds. Points from all strips are rasterized with cell size of  $[h_x, h_y, h_z]$  and a best-fitting plane is computed via principal component analysis (PCA) of all points within each cell (Fig. 5 and Fig. 6). Raster cells are classified as valid based on certain planarity conditions, e.g., if the smallest eigenvalue is below a given threshold. From this, we derive the plane center  $\tilde{\mathbf{c}}$  and (unit-length) normal  $\tilde{\mathbf{n}}$ . The plane center and normal are stored internally in the scanner coordinate system  $(\tilde{\mathbf{c}}_s, \tilde{\mathbf{n}}_s)$  and transformed into the local coordinate system according to the georeferencing equations

$$\begin{aligned} \tilde{\mathbf{c}} &= \mathbf{R}(t, \mathbf{x}_o)\mathbf{R}(\boldsymbol{\kappa})(\tilde{\mathbf{c}}_s + \mathbf{l}_s) + \mathbf{p}(t, \mathbf{x}_p), \\ \tilde{\mathbf{n}} &= \mathbf{R}(t, \mathbf{x}_o)\mathbf{R}(\boldsymbol{\kappa})\tilde{\mathbf{n}}_s. \end{aligned} \quad (3)$$

Every such *feature plane* is assigned the time  $t$  of the closest actual LiDAR point. The feature planes serve as LiDAR measurements  $(\tilde{\mathbf{c}}_s, \tilde{\mathbf{n}}_s)$  in the least-squares estimation. They are not actual measurements but rather each derived from multiple LiDAR points.

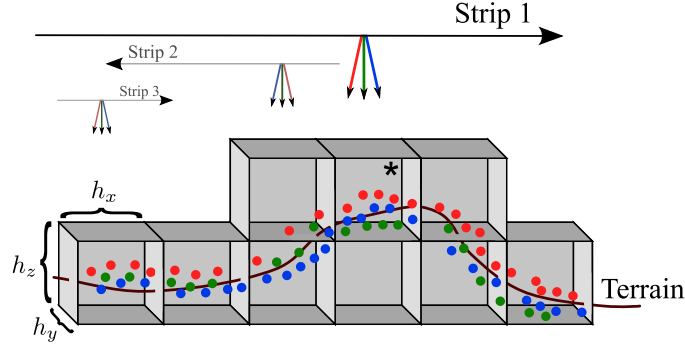


Fig. 5: All points are sorted into 3D raster cells but considered separately for each strip and view. Points are colored by view: forward (blue) / nadir (green) / backward (red)

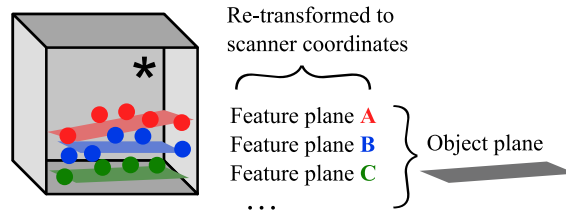


Fig. 6: For each strip and view, a feature plane is computed for the points in each raster cell via PCA. Feature planes in a cell are then associated with a corresponding object plane.

All feature planes in a given raster cell are aggregated into an *object plane*. This object plane corresponds to the actual physical plane, which is observed multiple times. Initially, each object plane is defined by a center  $\mathbf{c}^0$  and (unit-length) normal  $\mathbf{n}^0$ , as well as (unit-length) axes  $\mathbf{k}_1^0, \mathbf{k}_2^0$  so that  $\mathbf{n}^0 = \mathbf{k}_1^0 \times \mathbf{k}_2^0$ . For the least-squares adjustment, object planes are parametrized by an offset  $o$  and slopes  $z_1/z_2$ , with resulting center, axes and normal

$$\mathbf{c} := \mathbf{c}_0 + o \mathbf{n}^0, \quad \mathbf{k}_1 := \frac{\mathbf{k}_1^0 + z_1 \mathbf{n}^0}{\|\mathbf{k}_1^0 + z_1 \mathbf{n}^0\|}, \quad \mathbf{k}_2 := \frac{\mathbf{k}_2^0 + z_2 \mathbf{n}^0}{\|\mathbf{k}_2^0 + z_2 \mathbf{n}^0\|}, \quad \mathbf{n} := \mathbf{k}_1 \times \mathbf{k}_2. \quad (4)$$

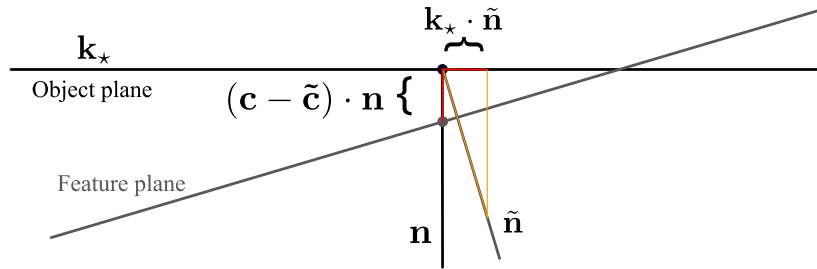


Fig. 7: Plane observation model: 1) Normal distance  $(\mathbf{c} - \tilde{\mathbf{c}}) \cdot \mathbf{n}$  from the object plane center  $\mathbf{c}$  to the feature plane center  $\tilde{\mathbf{c}}$ . 2) Axis distance  $\tilde{\mathbf{n}} \cdot \mathbf{k}_*$  for each object plane axis  $\mathbf{k}_1/\mathbf{k}_2$  and the feature plane normal  $\tilde{\mathbf{n}}$ .

The parameters  $(o, z_1, z_2)$ , and thus the object plane's position and orientation are optimized during the adjustment in order to minimize the normal and axis distances between object plane and all corresponding feature planes (Fig. 7). The plane measurement equation is given by

$$\begin{aligned}
 0 &= (\mathbf{c} - \tilde{\mathbf{c}}) \cdot \mathbf{n} + \epsilon_n, \\
 0 &= \tilde{\mathbf{n}} \cdot \mathbf{k}_1 + \epsilon_{k1}, \\
 0 &= \underbrace{\tilde{\mathbf{n}} \cdot \mathbf{k}_2}_{\text{constraint}} + \underbrace{\epsilon_{k2}}_{\text{error}}.
 \end{aligned} \tag{5}$$

Strictly speaking, this is not an explicit measurement, but rather a constraint. Note that the above equations implicitly depend on the platform trajectory and scanner calibration as in Equation (3). At this point, all errors occurring in the above measurement models are assumed Gaussian, uncorrelated and therefore independent. The relevant standard deviations for the Gaussian noise of the inertial measurements are determined in a static calibration. For the LiDAR correspondences, they are derived from the PCA results, taking into account LiDAR angular and ranging accuracy. The GNSS error standard deviations are internal estimates provided by the GNSS processing solution. However as argued above, the GNSS positions are not uncorrelated. Therefore, we now introduce a stochastic model for the time-correlated error component.

## 2.2 Stochastic Error Model for Loosely Coupled GNSS Integration

Assuming that the GNSS position measurements include not only white noise but also a time-correlated error term, the measurement equation can be extended to

$$\tilde{\mathbf{p}} = \mathbf{p}(t, \mathbf{x}_o) + \mathbf{R}(t, \mathbf{x}_o) \mathbf{l}_a + \underbrace{\mathbf{v}}_{\substack{\text{time-} \\ \text{correlated} \\ \text{errors}}} + \underbrace{\boldsymbol{\epsilon}}_{\substack{\text{uncorrelated} \\ \text{errors} \\ = \\ \text{measurement} \\ \text{noise}}}. \tag{6}$$

A simple, yet effective way of modelling time-correlated errors is a first-order Gauss-Markov (FOGM) process, defined recursively by

$$v_i = \underbrace{\exp\left(-\frac{t_i - t_{i-1}}{T}\right)}_{=:a} v_{i-1} + w_i, \quad w_i \sim N(0, \sigma_w^2). \tag{7}$$

This is an order 1 autoregressive process, where the autoregressive coefficient  $a$  is constrained by  $0 < a < 1$ . When the first sample  $v_0$  is also normally distributed with  $v_0 \sim N(0, \frac{\sigma_w^2}{1-a^2})$ , the process is stationary with covariance  $\mathbb{V}(v_i) = \frac{\sigma_w^2}{1-a^2}$  and zero mean  $\mathbb{E}(v_i) = 0$ .

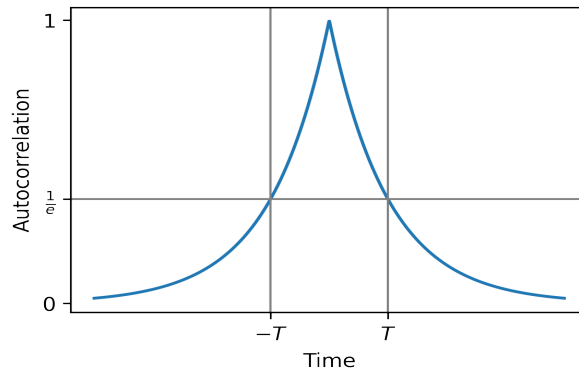


Fig. 8: Autocorrelation of first-order Gauss-Markov process  $v$

For each coordinate system axis of the local coordinate system, we model a separate Gauss Markov process. Each process is defined by two parameters, its process noise standard deviation and its correlation time. The process noise variance  $\sigma_w^2$  controls the amount of stochastic noise added in each time step. While the measurement noise is uncorrelated, each  $v_i$  is a weighted sum of the past process noise realizations and therefore correlated in time. The correlation time  $T$  is a measure of how long it takes the process to become uncorrelated with itself; specifically,  $T$  refers to when the autocorrelation drops below  $1/e$ . If  $T \rightarrow 0$ , the additive white-noise error model is recovered, while for  $T \rightarrow \infty$  the FOGM turns into a random walk. For errors in GNSS positioning, time correlation of around 10-30 seconds may be expected (from own experiments), but this can go up to several 100 seconds in the presence of multi-path effects (HAN & RIZOS 2000). Since in most instances, the applicable parameters of the Gauss-Markov process are not known, we propose to adapt the Cochrane-Orcutt estimator (COCHRANE & ORCUTT 1949) from linear least-squares with time-correlated errors to our specific non-linear least squares setting.

In our model, the unknown time-correlation applies only to the GNSS errors. The time-correlated component  $\mathbf{v}$  is explicitly modelled by the parameters  $\mathbf{b}_g$ , assumed to follow a first-order Gauss-Markov process. The Cochrane-Orcutt estimator is adapted in the following way:

1. Estimate model parameters with uncorrelated GNSS measurement model.
2. Obtain an estimate of the process correlation time  $T$  and standard deviation  $\sigma$  by fitting a first-order Gauss Markov process to the residuals

$$\mathbf{r} := \tilde{\mathbf{p}} - \mathbf{p}(t, \mathbf{x}_o) + \mathbf{R}(t, \mathbf{x}_o)\mathbf{l}_a = \mathbf{v} + \boldsymbol{\varepsilon}. \quad (8)$$

Per assumption, this is an FOGM process with additive measurement noise. For observation  $i$ , the process model has the form

$$\underbrace{\begin{pmatrix} r_x \\ r_y \\ r_z \end{pmatrix}_i}_{\text{GNSS residuals}} = \underbrace{\begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix}_i}_{\text{Gauss-Markov process}} + \underbrace{\begin{pmatrix} \varepsilon_x \\ \varepsilon_y \\ \varepsilon_z \end{pmatrix}_i}_{\text{measurement noise}}, \quad (9)$$

$$\begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix}_i = \begin{pmatrix} a_x & 0 & 0 \\ 0 & a_y & 0 \\ 0 & 0 & a_z \end{pmatrix} \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix}_{i-1} + \underbrace{\begin{pmatrix} w_x \\ w_y \\ w_z \end{pmatrix}_i}_{\text{process noise}}.$$

The coefficients  $(a_x, a_y, a_z)$  and hence the correlation times, as well as process noise variances and measurement noise variances can be estimated using standard state-space methods from time-series analysis (DURBIN & KOOPMAN 2012).

3. Estimate model parameters with time-correlated GNSS errors via non-linear least-squares.

Note that in the above, the measurement noise is assumed Gaussian, independent and identically distributed with no cross-correlation between the axes. This is not typically the case for the covariance matrices of the GNSS position measurements which are provided by the GNSS processing solution, as these include some correlations between the x-, y- and z-axis. This assumption is a limitation due to the implementation of the time-series estimation; the full covariance matrix of the GNSS position measurement is used in the actual adjustment.

### 3 Airborne Laser Scanning Case Study

#### 3.1 Equipment and Dataset

The data for this case study was acquired with a RIEGL VUX-160<sup>23</sup> and an AP+50 GNSS/INS navigation system mounted on a Cessna 206 over Retz, a small town in Lower Austria. The flight trajectory is shown in Fig. 10. The flying altitude was 550 m above ground level with a ground speed of 57 m/s and a pulse repetition rate of 600 kHz, resulting in a point density per strip of 7 points per square meter.

The RIEGL VUX-160<sup>23</sup> is an airborne laser scanner with a mass of 2.65 kg, a field of view of 100 degrees, a scan speed of up to 400 lines/sec and a pulse repetition rate of up to 2.4 MHz. It is designed for corridor mapping applications with high point density. The instrument is based on time-of flight measurements at 10 mm accuracy, 5 mm precision, and a measurement range of up to 1800 m. It is able to capture up to 32 targets per pulse and thus offers excellent multi-target capabilities, which is especially important for power line surveying and forestry. The RIEGL VUX-160<sup>23</sup> offers a sophisticated scan pattern consisting of scan lines with periodically changing directions (Fig. 3). The scan directions in the center of the scan lines change consecutively from strictly nadir, to +10 degrees forward, and to -10 degrees backward. This scan pattern provides an almost complete 3D data set, as also vertical surfaces like the facades of buildings and objects (e.g. towers, masts, and poles) are accurately sampled by laser range measurements (Fig. 9). In addition, the nadir direction enables reliable data acquisition down to the bottom of narrow canyons.

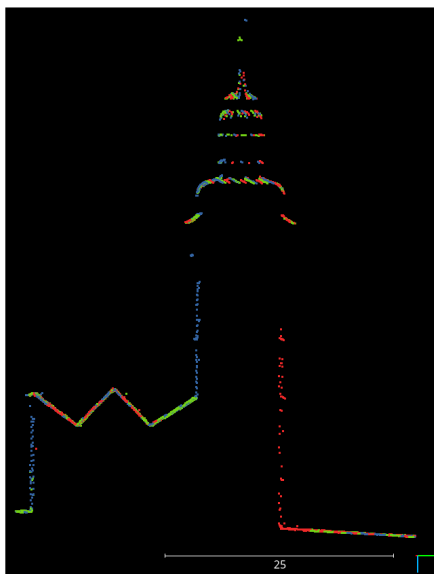


Fig. 9: Retz chapel. The points are colored according to the three different view directions indicated in Fig. 4

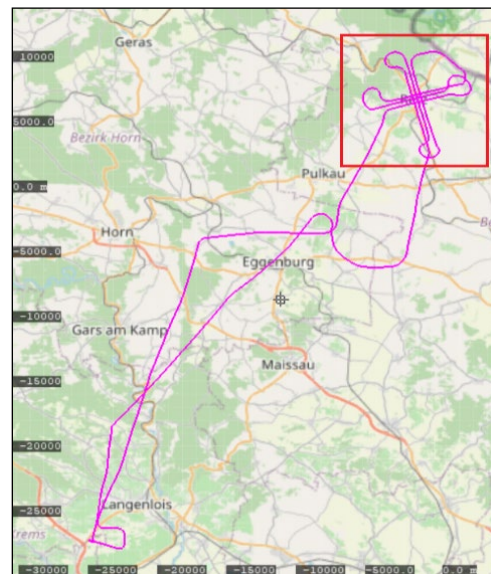


Fig. 10: Flight trajectory with 2x3 flight strips in a cross pattern

#### 3.2 Results and Comparison

For comparison and to motivate the methodology, we compute a separate strip adjustment solution with the OPALS software (PFEIFER et al. 2014), using the approach developed by GLIRA et al.

2016. In this adjustment, laser scanner calibration (lever arm and boresight angles) are estimated as part of the adjustment, and the initial trajectory from GNSS/INS is corrected by estimating a constant offset in position and orientation per strip. Correspondences in the laser point cloud are established iteratively via a closest point algorithm, while the underlying GNSS and INS measurements are not considered.

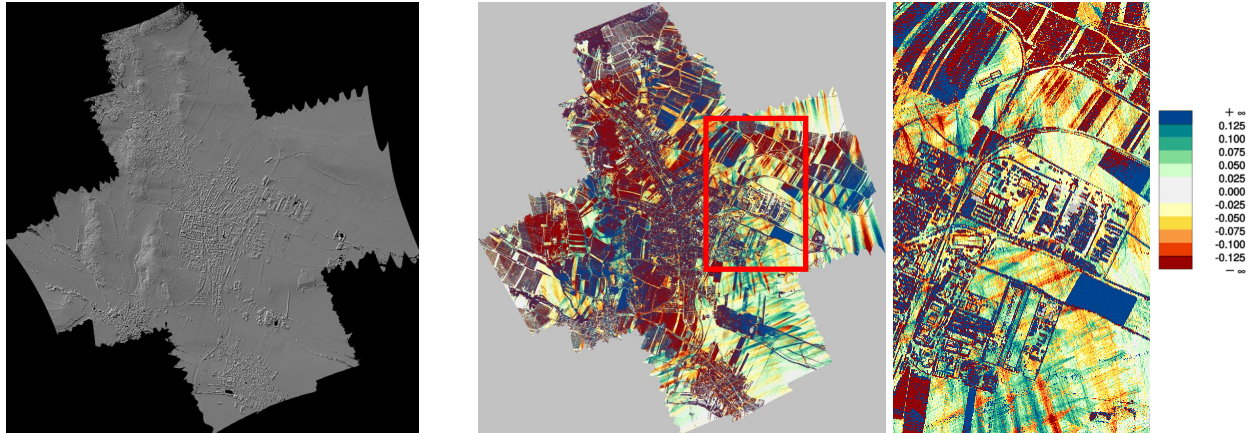


Fig. 11: DSM of all strips and views    Fig. 12: Signed maximum strip difference of all strips

The study area contains a large number of fields and other vegetation, which will show up as large discrepancies in the strip differences (Fig. 12) due to differing viewing directions. For a clearer presentation, we will focus on the area marked in red. The strip differences show systematic errors up to 15 cm, presumably stemming in a large part from an inaccurate trajectory. These errors are on a timescale of a few seconds or less, making them impossible to correct with a constant per-strip offset. One approach would now be to allow not only an offset, but rather a time-varying correction using splines. This then leads to the problem of choosing the number of spline parameters or equivalently, the spacing of the spline knots: If the spacing is too large, the errors will not be fully correctable. A smaller spacing however may quickly lead to overfitting and deformation of the trajectory and in consequence the point cloud. This leads us to the approach detailed above: Keep the measurements from GNSS and INS and integrate them with LiDAR correspondences.

The GNSS/INS/LiDAR-integration is carried out as described in Section 2. Inertial measurements are available with 200 Hz, GNSS positions with 1 Hz. The frequency and number of the LiDAR plane correspondences depends on the rasterization parameters. Here, a cell size of 2.5 m was used, followed by a 4x sub-sampling of all planes and a 24x sub-sampling of all horizontal planes. The forward/nadir/backward viewing angles allow extraction of correspondences even on short timescales (within a strip) and with very steep slopes (e.g., facades). Comparing Fig. 13 and Fig. 14 there are two major differences: The overall number of extracted objects is much lower with only the nadir view, because in-strip overlaps of the different views are not considered. However, in addition to an overall reduction in the number of extracted object planes, there are almost no vertical planes left at all, as these are rarely seen with a nadir viewing angle.



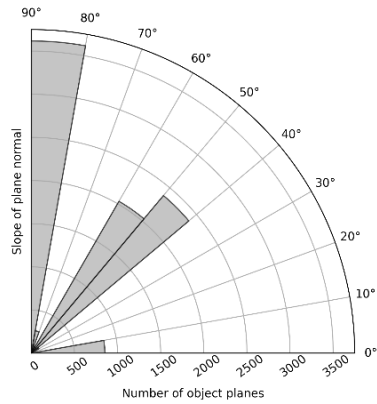


Fig. 13: Number of extracted object planes per 10-degree interval of the slope of the plane normal, for all views of all strips

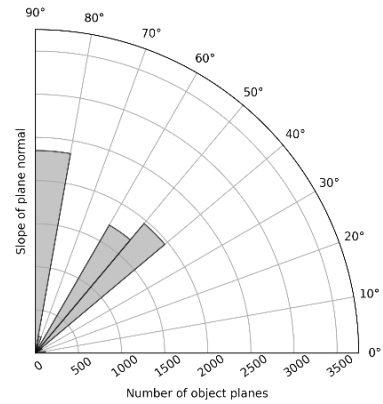


Fig. 14: Number of extracted object planes per 10-degree interval of the slope of the plane normal, for the nadir view of all strips

Analogously to Fig. 12, the strip differences after the batch least-squares adjustment using GNSS positions, inertial measurements and LiDAR plane correspondences are shown in Figure 15. However, the errors have different characteristics compared to those in Fig. 12: There are large-scale differences of up to 15 cm, while the smaller-scale difference patterns have mostly disappeared. While it is hard to pinpoint the error source exactly due to the complexity of the model, discrepancies that are constant over wider areas point towards an inaccurate platform position. As shown in Fig. 17, the residuals for the GNSS position observations increase drastically within all strips. In order to achieve a consistent point cloud, the modelled antenna position has to diverge significantly from the measured GNSS position.

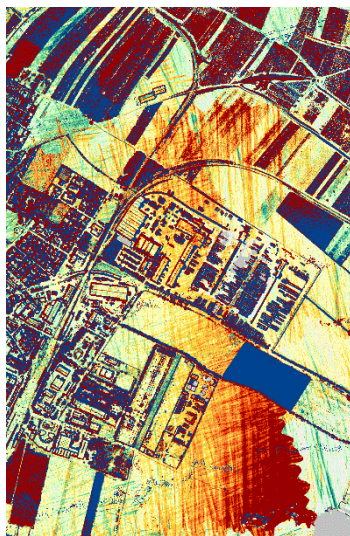


Fig. 15: Strip differences of GNSS/INS/LiDAR-adjustment with uncorrelated GNSS positions

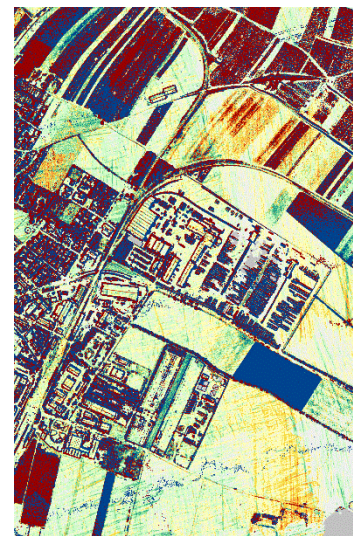


Fig. 16: Strip differences after GNSS/INS/LiDAR-adjustment with time-correlated GNSS position errors.

The differences between the GNSS position measurements and the modelled antenna position (Fig. 17) have a systematic time-dependence. This is contrary to the stochastic assumption of measurements with independent zero-mean noise, resulting in the strip differences seen above. The discrepancies between the model and the GNSS measurements are largest in the center of each strip, where LiDAR correspondences with the cross-track strips are present.

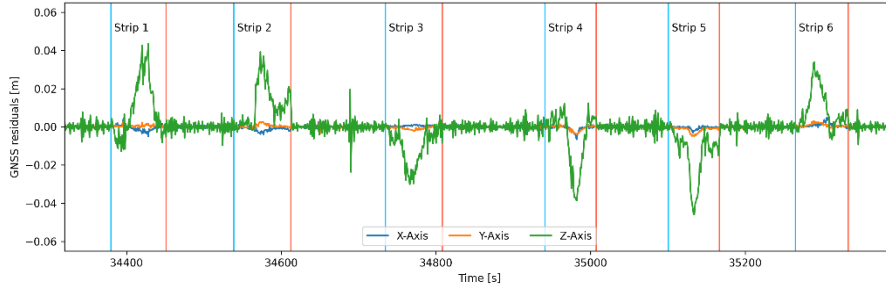


Fig. 17: GNSS residuals without time-correlated error model

Table 2 shows the estimated correlation time and process noise standard deviations of the GNSS position residuals. The correlation time is 20-35 seconds depending on the axis. The process noise standard deviation is 10x higher for the z-axis than for the x- and y-axis. This partially due to the LiDAR measurements, as the majority of the correspondences are horizontal planes, which constrain the trajectory only in the height component. The increased process noise standard deviation in the z-axis may also suggest the actual error process is more complicated than the modelled FOGM process. Nevertheless, for representing the time-correlated characteristics of the residuals in Fig. 17, the FOGM model is a substantial improvement compared to assuming independent measurements.

Tab. 2: Estimated Gauss-Markov process parameters

	x-Axis	y-Axis	z-Axis
Correlation time $T$ [s]	21.14531	26.97508	34.87152
Process noise standard deviation <sup>1</sup> $\sigma_w$ [m]	0.00021	0.00019	0.00227

With the estimated process parameters, we can now model the time-correlated GNSS bias within the least-squares adjustment, while all other parameters remain the same. The resulting point cloud shows drastically reduced differences compared to before (Fig. 16). Judging from the comparatively homogenous strip differences of the rooftops on the lower left side of Fig. 16, there seems to be no significant lateral error. The magnitude of the estimated bias  $\mathbf{b}_g$  is around 4-5cm for the z-axis, and the GNSS residuals now show no obvious time-correlation (Fig. 18). Some discrepancies between different strips and between views within a strip do remain, the cause of which will be investigated in future work.

<sup>1</sup> The noise standard deviations are of course only applicable for our time sampling frequency of 1Hz (or equivalently, a time step size of 1s) and would have to be converted for other sampling frequencies.

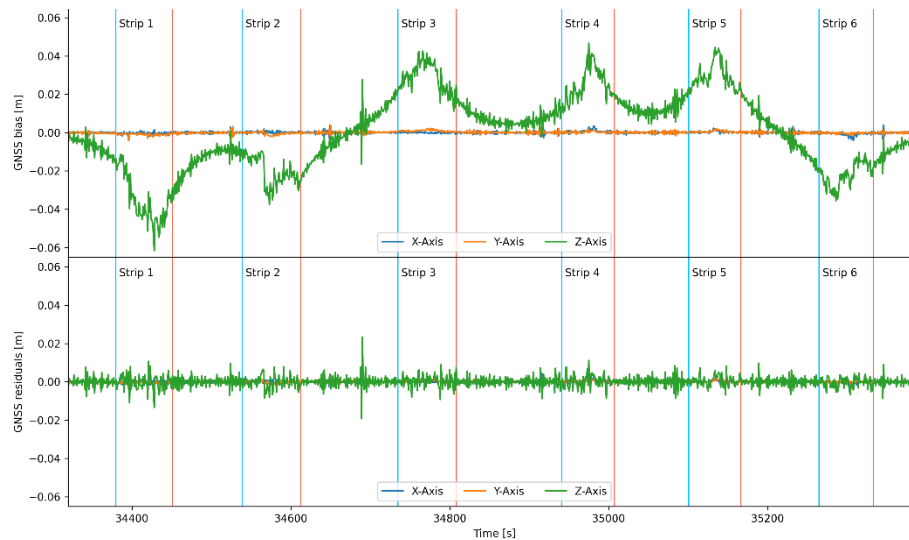


Fig. 18: GNSS errors and residuals with time-correlated error model

## 4 Conclusion

In this contribution, we presented a method for integrating GNSS, INS and LiDAR data, which takes into account the time-correlation of the GNSS position measurements. The approach was validated with an ALS dataset where it achieved reduced strip differences compared to a classic strip adjustment approach using a constant per-strip trajectory correction and to the GNSS/INS/LiDAR-integrated adjustment without explicit modelling of the GNSS errors' time-correlation. However, this does not conclusively disprove the existence of strip deformations, which might occur with a more flexible spline-based strip adjustment. In future work, it would be prudent to perform a systematic comparison of the results not only in terms of internal consistency (i.e., strip differences), but also with accurate reference data in order to rule out these systematic deformations in the point cloud.

## 5 Bibliography

- AGARWAL, S., MIERLE, K. & The Ceres Solver Team, 2022: Ceres Solver, <https://github.com/ceres-solver/ceres-solver>.
- COCHRANE, D. & ORCUTT, G.H., 1949: Application of Least Squares Regression to Relationships Containing Auto-Correlated Error Terms. *Journal of the American Statistical Association*, **44**(245), 32-61.
- DURBIN, J. & KOOPMAN, S.J., 2012: *Time Series Analysis by State Space Methods*. 2nd ed., Oxford University Press, Oxford.
- FARRELL, J.A., SILVA, F.O., RAHMAN, F. & WENDEL, J., 2022: Inertial Measurement Unit Error Modeling Tutorial: Inertial Navigation System State Estimation with Real-Time Sensor Calibration. *IEEE Control Systems Magazine*, **42**(6), 40-66.
- FILIN, S. & VOSSELMAN, G., 2004: Adjustment of Airborne Laser Altimetry Strips. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, 258-263.

- GLIRA, P., PFEIFER, N. & MANDLBURGER, G., 2016: Rigorous Strip Adjustment of UAV-based Laserscanning Data Including Time-Dependent Correction of Trajectory Errors. *Photogrammetric Engineering & Remote Sensing*, **82**(12), 945-954.
- GROVES, P.D., 2013: *Principles of GNSS, Inertial, and Multisensor Integrated Navigation Systems*. 2nd ed., Artech House, Boston.
- HAN, S. & RIZOS, C., 2000: GPS Multipath Mitigation Using FIR Filters. *Survey Review*, **35**(277), 487-498.
- KAGER, H., 2004: Discrepancies Between Overlapping Laser Scanner Strips-Simultaneous Fitting of Aerial Laser Scanner Strips. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, **35**, 555-560.
- KIM, M.-J., KIM, M.-S. & SHIN, S.Y., 1995: A General Construction Scheme for Unit Quaternion Curves with Simple High Order Derivatives. *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques - SIGGRAPH '95*, ACM Press, 369-376.
- NIU, X., WU, J. & ZHANG, Q., 2018: Research on Measurement Error Model of GNSS/INS Integration Based on Consistency Analysis. *GyroscoPy and Navigation*, **9**(4), 243-254.
- PFEIFER, N., MANDLBURGER, G., OTEPKA, J. & KAREL, W., 2014: OPALS – A Framework for Airborne Laser Scanning Data Analysis. *Computers, Environment and Urban Systems*, **45**, 125-136.
- PÖPPL, F., NEUNER, H., MANDLBURGER, G. & PFEIFER, N., 2023: Integrated Trajectory Estimation for 3D Kinematic Mapping with GNSS, INS and Imaging Sensors: A Framework and Review. *ISPRS Journal of Photogrammetry and Remote Sensing*, **196**, 287-305.
- TAKASU, T. & YASUDA, A., 2009: Development of the Low-Cost RTK-GPS Receiver with an Open Source Program Package RTKLIB. *International Symposium on GPS/GNSS*, Jeju, Korea, **7**.

# Assessing the Potential of High-Resolution Satellite Constellations for Agricultural Monitoring

SAMUEL WILDHABER<sup>1,2</sup>, LUKAS V. GRAF<sup>1,2</sup> & HELGE AASEN<sup>2</sup>

*Abstract: Planet-Labs' SuperDove fleet shows potential for agricultural monitoring through high spatial (ground sampling distance <5m) and temporal resolution (multiple scenes per day). We retrieved widely used vegetation indices (VIs) and the leaf area index (LAI) from radiative transfer model inversion for a winter wheat parcel over an entire growing season and compared the data to ground samples and Sentinel-2 imagery. LAI was retrieved with an RMSE of 0.91 m<sup>2</sup>m<sup>-2</sup> (R<sup>2</sup>=0.87), which was similar to the retrieval performance of Sentinel-2 (RMSE=0.87 m<sup>2</sup>m<sup>-2</sup>, R<sup>2</sup>=0.87). The SuperDove imagery revealed a higher level of spatial detail than Sentinel-2 and replicated spatio-temporal field heterogeneity. Still, issues regarding radiometric stability, spatial co-registration accuracy and quality masks need to be addressed to use the SuperDove imagery for monitoring activities.*

## 1 Introduction

Capturing plant growth conditions at landscape scale offers unprecedented opportunities to improve understanding of plant-environment interactions, which can help increasing sustainability of agricultural production. Remote sensing platforms such as Landsat or Sentinel-2 have already demonstrated their potential for large-scale applications but also at the sub-field level (WEISS et al. 2020). However, in agricultural landscapes with small field sizes and highly diversified spatial patterns as they are widespread in Switzerland, the spatial resolution of the platforms (30 to 10m) is not sufficient to cover all field parcels and small-scaled spatial patterns. Until now, satellite constellations providing high spatial (< 5m) and temporal resolution (up to one day) were limited with respect to spectral information, i.e., the number of spectral channels and their quality was lower compared to Sentinel-2 or Landsat. With the advent of the most recent Super-Dove cube satellites from Planet-Labs, multi-spectral data with eight reflective channels at 3m spatial resolution are available, including the red edge spectral region, which is important for vegetation studies. This makes the use of the Super-Dove constellation potentially interesting for agroecological research.

In the present work, we evaluate the potential of the SuperDove constellation for agricultural monitoring. We use vegetation indices such as the Normalized Difference Vegetation Index (NDVI) as well as the Leaf Area Index (LAI) from the inversion of the radiative transfer model PROSAIL (JACQUEMOUD et al. 2009) over an entire growing season. We compare results to in-situ reference and Sentinel-2 data to provide a comparison to established remote sensing-based agricultural monitoring systems. Based on this, we discuss the advantages and disadvantages of

---

<sup>1</sup> ETH Zürich, Kulturpflanzenwissenschaften, Universitätstrasse 2, CH-8092 Zürich, Schweiz, E-Mail: lukasvalentin.graf@usys.ethz.ch

<sup>2</sup> Agroscope Reckenholz, Fernerkundung der Agrarlandschaft, Reckenholzstrasse 191, CH-8046 Zürich, Schweiz, E-Mail: helge.aasen@agroscope.admin.ch

the SuperDove constellation and present some challenges with regard to a productive use of the data.

## 2 Data and Methods

### 2.1 Study Area and in-situ reference data

To evaluate the suitability of the Super-Dove satellites for agricultural monitoring, a winter wheat field parcel (2.04ha) located close to Lindau-Eschikon, Switzerland (Lat. 47.45, Lon. 8.69) was chosen during the 2022 growing season. The area receives an annual precipitation total of 1200mm with an annual mean air temperature of 10° C (2011-2021 reference period). Soils in the study area are slightly alkaline (pH between 7.2 and 7.8) and are considered loamy with clay content of 20 to 30% and a moderate humus content of 3.0 to 3.6%. The parcel is moderately sloped, which is typical for this region of Switzerland. The field was managed conventionally according to Swiss standards. Mineral fertilizer was applied three times during the vegetation period, organic fertilizer only before sowing in October 2021.

Non-destructive LAI measurements were collected every two to four days at 29 pre-defined locations from April to end of June using a LAI-2200C Plant Canopy Analyzer from LI-COR Biosciences. BBCH ratings were accomplished by eye to record phenological development. This resulted in a total of 942 LAI measurements and BBCH ratings. Figure 1 shows the location of the field parcel in Switzerland (a) and the locations where LAI was measured within the parcel boundaries with underlying Sentinel-2 (Fig. 1b) and SuperDove (Fig. 1c) true-color RGB imagery acquired on May 14<sup>th</sup> 2022.

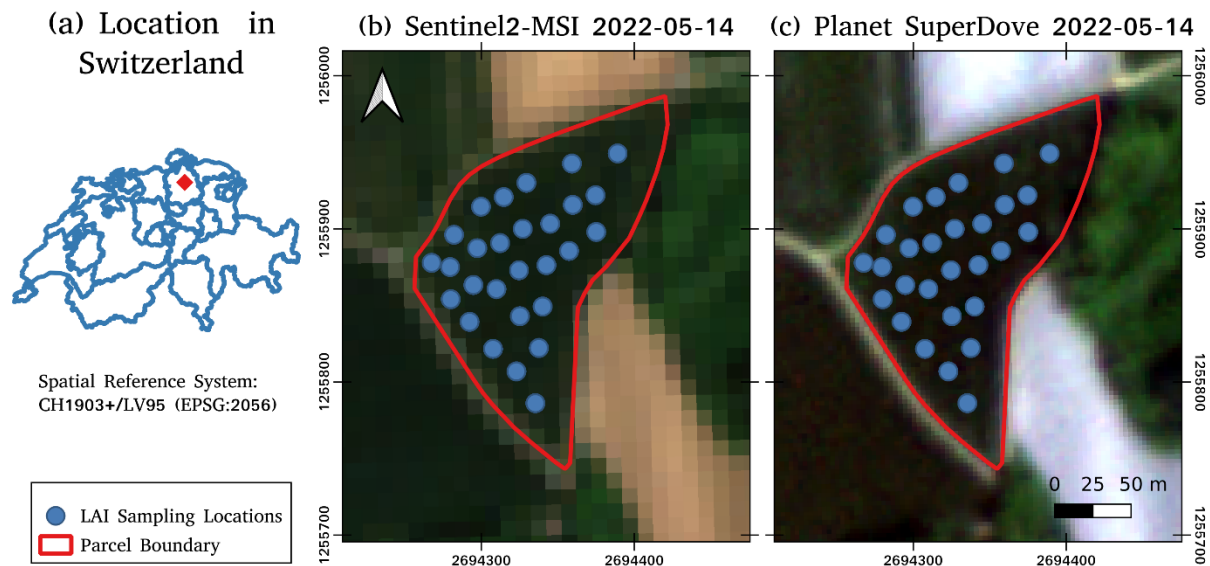


Fig. 1: Maps showing the location of the study area in Switzerland as red triangle in (a) and the locations where LAI samples were taken with underlying Sentinel-2 (b) and Planet SuperDove (c) true-color imagery

## 2.2 Satellite Data

Satellite data acquisition, preprocessing and extraction was undertaken using the open-source Python “Earth Observation data analysis library” (EOdal, GRAF et al. 2022). Both, Sentinel-2 and PlanetScope SuperDove data were acquired in UTM coordinates (Zone 32N).

### 2.2.1 Planet SuperDove

Planet SuperDove 8-band imagery in 3m spatial resolution was downloaded using the Planet Orders API. The data was corrected for atmospheric effects, orthorectified, and converted to surface reflectance factors by Planet and delivered alongside a usable data mask (UDM2-layer) providing clear pixel probabilities. All scenes were co-registered using the “arosics” library (SCHEFFLER et al. 2017). For reference, a clear SuperDove scene from March 10, 2022, was chosen. In total, 53 scenes with a scene-wide cloud cover smaller than 50% were available for analysis. From these, we excluded nine scenes from analysis after visual inspection because of undetected clouds and other distortions.

### 2.2.2 Sentinel2-MSI

Atmospherically corrected Sentinel2-MSI data was downloaded from CREODIAS including a scene-classification layer to filter out cloud or shadow pixels (tiling grid 32TMT). All 20m bands were resampled to a spatial resolution of 10m. We disregarded the 60m bands for further analysis. In total, 17 scenes with a scene-wide cloud cover smaller than 50% were available for analysis.

### 2.2.3 Satellite Data Processing

Clouds and shadows were masked in both satellite datasets using the provided clear pixel masks. We then calculated three widely used spectral vegetation indices sensitive to canopy greenness (NDVI), nitrogen content (Normalized Difference Red-Edge, NDRE) and chlorophyll content (CI\_GREEN). Furthermore, we used the radiative transfer model PROSAIL to derive the LAI using a lookup-table based retrieval approach. Lookup-tables were generated by running PROSAIL 50 000 times in forward mode using a distribution of leaf and canopy parameters proposed by DANNER et al. (2021). For inversion, we used the median of the 5000 lookup-table entries in terms of the smallest spectral root mean squared error (RMSE) between PROSAIL-simulated and satellite-observed spectra.

Satellite derived LAI values were compared to in-situ LAI measurements with a maximum temporal offset of one day. We calculated common error metrics such as RMSE, normalized absolute median deviation (NMAD) and the coefficient of correlation (Pearson’s R-Square). For the validation of the SuperDove-derived LAI, we used all available sampling points whereas the Sentinel-2 data was validated on a reduced data set due to constraints regarding the spatial and temporal resolution.

To provide a direct comparison between Sentinel-2 and SuperDove, we aggregated the SuperDove pixels into the spatial resolution of the Sentinel-2 pixels: For this, we used the mean of all SuperDove pixels that are fully contained in a Sentinel-2 pixel to allow for a pixel-based comparison of vegetation index and LAI values.

### 3 Results

#### 3.1 Leaf Area Index Retrieval

Figure 2 shows the results of comparing Sentinel-2 (Fig. 2a) and SuperDove-derived LAI (Fig. 2b) against ground reference data with a maximum temporal gap between satellite overpass and ground sampling of one day. Due to the coarser spatial and temporal resolution of Sentinel-2 the amount of ground reference data points ( $N=28$ ) matching the satellite overpasses is significantly lower than in the case of the SuperDove constellation ( $N=444$ ). In both cases, high LAI values were slightly underestimated resulting in an overall RMSE of  $0.87 \text{ m}^2 \text{ m}^{-2}$  and  $0.91 \text{ m}^2 \text{ m}^{-2}$  for Sentinel-2 and SuperDove, respectively. Still, the satellite data explained 87% of the variance in the reference data in both cases. NMAD was high for Sentinel-2 ( $0.34 \text{ m}^2 \text{ m}^{-2}$ ) and close to zero ( $0.01 \text{ m}^2 \text{ m}^{-2}$ ) in the case of Planet SuperDove. Although the validation results of Sentinel-2 and Planet SuperDove are not directly comparable both systems reproduce overall LAI values and temporal development with accuracies comparable to findings in the literature (e.g. DANNER et al. 2021). With regard to phenology, LAI relative retrieval accuracy is lowest during tillering (BBCH stages 20 to 29) with relative errors up to  $>50\%$  and during flowering (BBCH 61 to 69). Smaller relative errors and higher  $R^2$  values can be found during the stem elongation phase (BBCH 31 till 59) where errors drop down to 10% and  $R^2$  reaches up to  $>0.6$ . These findings are similar for Sentinel-2 and Planet SuperDove. We did not assess accuracy during senescence because of the difficulty to separate green from brown LAI and the inability of the PROSAIL RTM to account for non-photosynthetically active foliage area.

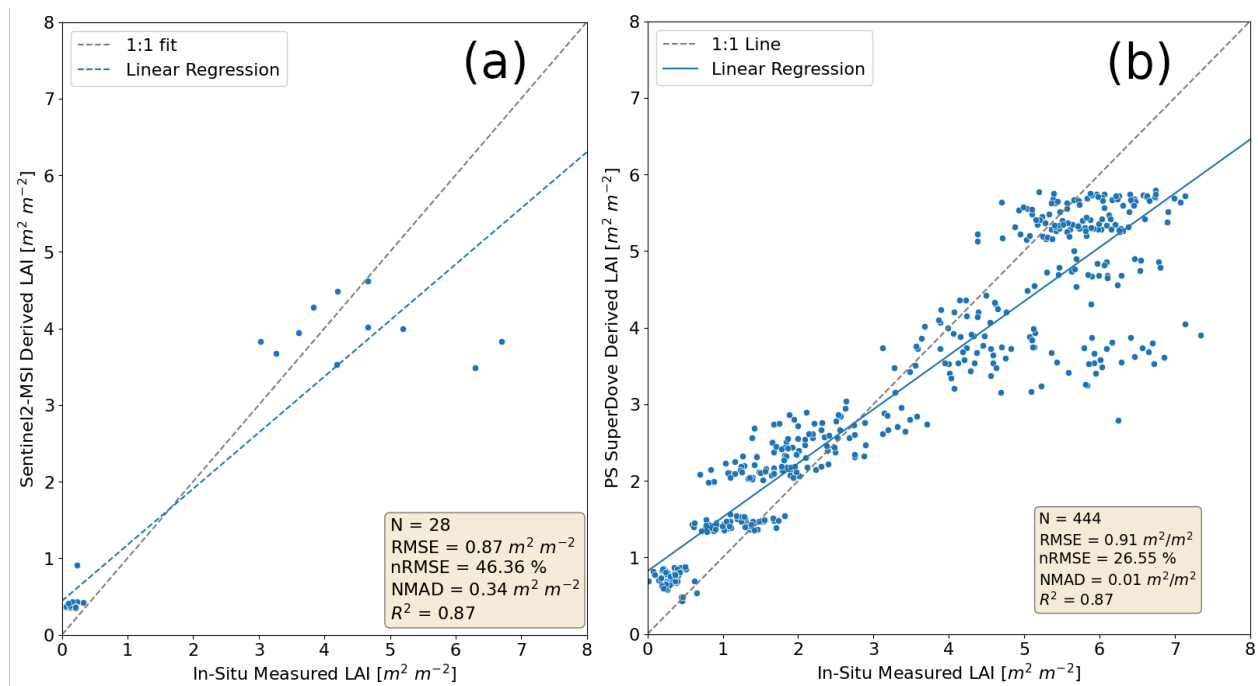


Fig.2: Validation of Sentinel-2 (a) and SuperDove (b) derived LAI values against ground reference data



### 3.2 Pixel-based comparison

Fig. 3 shows maps and pixel-based comparison of NDVI, NDRE, CI-GREEN, and LAI derived from Sentinel-2 and Planet SuperDove imagery on March 25 2022 before the second nitrogen fertilizer application. The plants were in tillering stage (BBCH 21 to 29). The amount of spatial detail in the SuperDove imagery is clearly higher than in the Sentinel-2 derived maps. While LAI shows a relatively strong agreement on the level of the Sentinel-2 10m pixels ( $R^2=0.46$ , relative error 11.7%), the systems disagree in the vegetation indices with lowest values in the case of CI-GREEN ( $R^2=0.14$ , relative error 38.4%) and NDRE ( $R^2=0.15$ , relative error 15%). SuperDove derived values tend to be higher than those from Sentinel-2. In the case of NDVI the relative error is small (10.3%) but  $R^2$  is only about 0.27 meaning that a large part of the variance in the Planet imagery is not explained by Sentinel-2.

Still in tillering phase, before the third nitrogen application, on April 19 2022 (Fig. 4), the agreement between Sentinel-2 and SuperDove is higher especially for NDVI ( $R^2=0.71$ , relative error 2.1%) but also NDRE and CI-GREEN. The SuperDove derived maps show a spot of relatively low values in the upper right part of the field parcel, which is hardly visible in the Sentinel-2 maps. This area of approximately 15 by 15 meters corresponds to a wet spot in the field as confirmed by the local farmer.

## 4 Discussion

### 4.1 Potentials

Both Sentinel-2 and Planet SuperDove can map winter wheat growth dynamics and allow accurate estimation of LAI from RTM inversion (Fig. 2). The SuperDove data, due to its superior spatial and temporal resolution, allows capturing small-scale details and field heterogeneities better compared to Sentinel-2 data. In particular, in the small fields of Switzerland with complex geometric shapes, the area covered by SuperDove is much larger, since a smaller boundary area must be buffered to avoid spectrally mixed pixel effects. The high spatial coverage thus allows establishing agricultural monitoring and decision systems - for example for nitrogen fertilization planning – that consider in-field heterogeneity (Fig. 3 and 4). Due to the high temporal coverage of the Planet data, growth dynamics and changes are captured, as sufficient cloud-free images are available even at high levels of cloud cover in the spring months. In this work, 42 Planet SuperDove scenes were available, compared to 17 Sentinel-2 images.

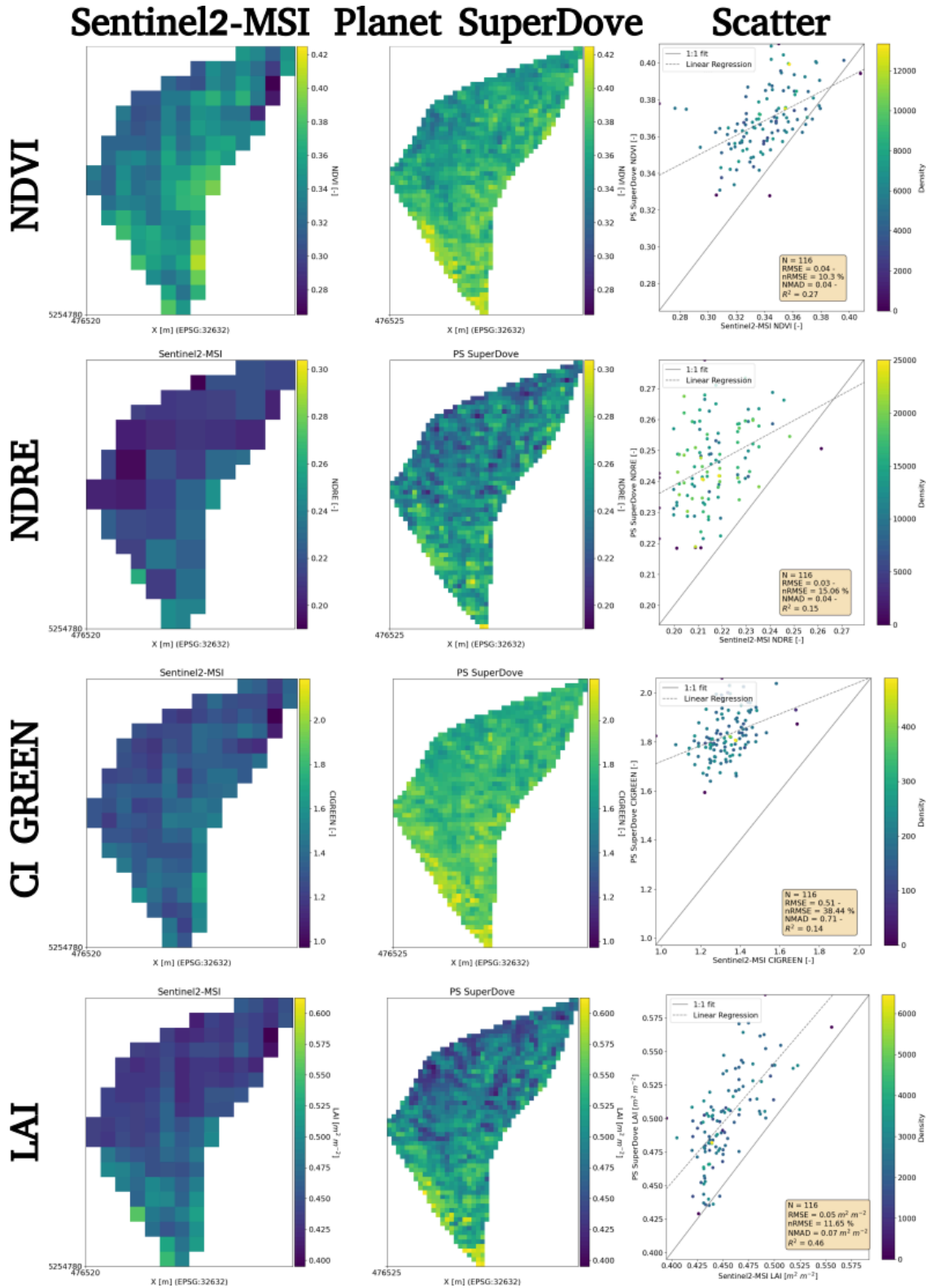


Fig. 3: Comparison of NDVI, NDRE, CI-GREEN and LAI between Sentinel-2 (left column), SuperDove (middle column) on a per-pixel level (right column) using the Sentinel-2 10m pixels for March 25 2022 (before 2nd nitrogen fertilization)

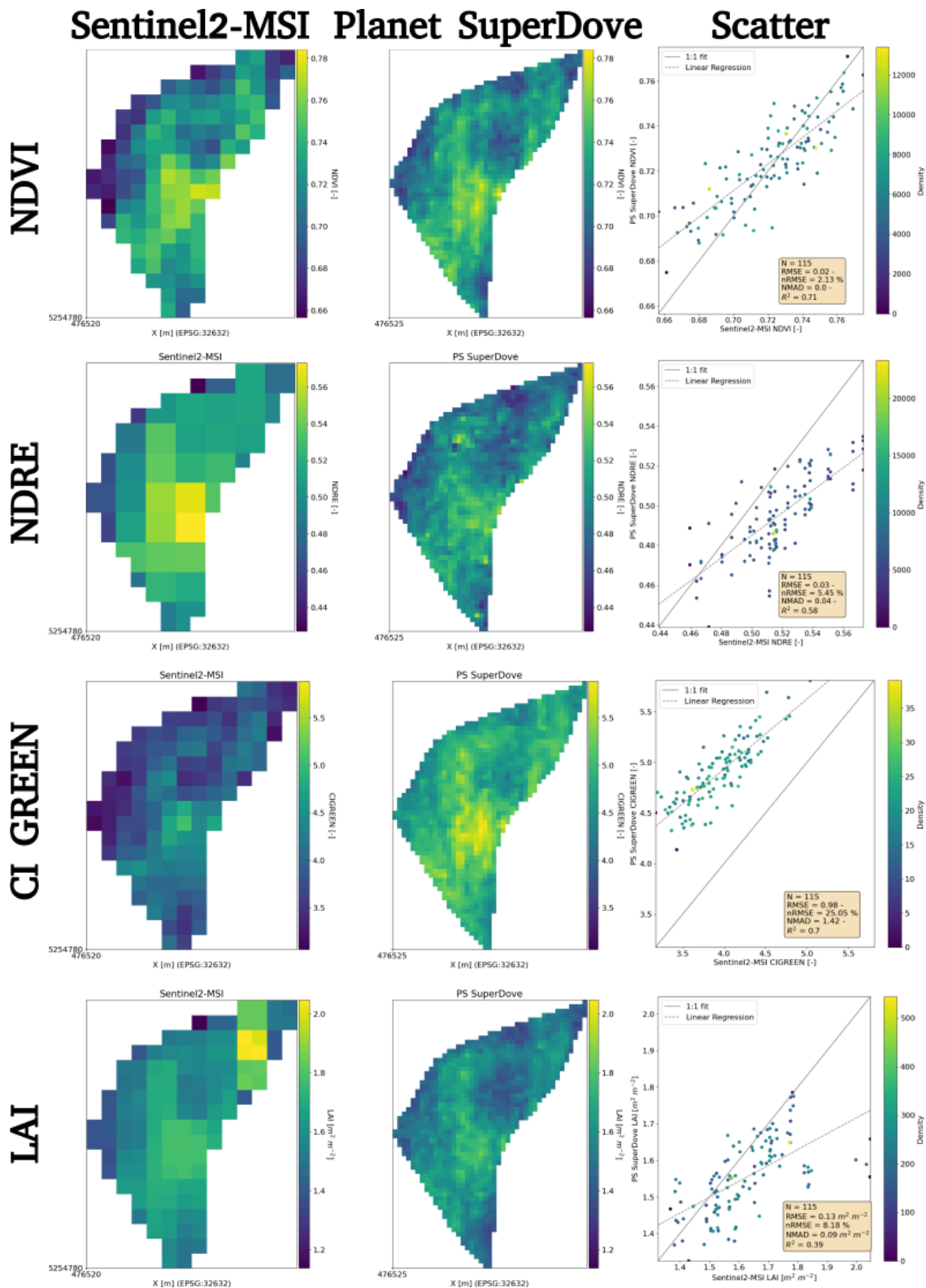


Fig. 4: Comparison of NDVI, NDRE, CI-GREEN and LAI between Sentinel-2 (left column), SuperDove (middle column) on a per-pixel level (right column) using the Sentinel-2 10m pixels for April 19 2022 (before 3rd nitrogen fertilization)

## 4.2 Challenges

Fig. 3 and 4 show only partial agreement between SuperDove and Sentinel-2 vegetation indices and GLAI. We attribute this observation partly to spatial co-registration errors, radiometric stability issues and the quality of the clear pixel mask.

### 4.2.1 Spatial Co-Registration Error

Accurate spatial co-registration of data is essential for recording vegetation dynamics in agricultural monitoring. While the European Space Agency (ESA) guarantees a co-registration error of <10m since the last major Payload Data Ground Segment (PDGS) baseline update to version N0400, the error in the SuperDove data is specified as less than 10m RMSE at the 90th percentile according to Planet Labs. Internal tests on a spatially invariant target (Zurich airport tarmac), however, showed that the offset of the scenes from each other was larger than specified. Subsequent co-registration with the “arosics” library still resulted in uncorrected spatial offsets.

### 4.2.2 Radiometric Stability

The fleet of SuperDove cube satellites enables high spatial and temporal resolution. However, the instruments are only conditionally radiometrically harmonized (TU et al. 2022). Although Sentinel-2A and -2B also differ slightly radiometrically, with only two sensors the number is significantly smaller than for Planet SuperDove, where data from 32 different satellites were included for the period under consideration. The high temporal resolution also means that in some cases more than one SuperDove image per day is available resulting in differences in the illumination geometry and bi-directional reflectance (BRDF) effects. We ignored all but the first of several daily scenes, but this is not necessarily the way to go. Radiometric harmonization and BRDF correction are therefore essential for operational use of Planet-Scope data for agricultural monitoring.

### 4.2.3 Quality of the Clear Pixel Mask

As described in 2.2.1, we sorted out nine SuperDove scenes (~16% of the downloaded scenes) manually. This shows that the clear pixel mask supplied with the scenes is not yet sufficiently reliable to grant automatic selection of the scenes. In particular, we found disturbances due to undetected clouds and shadow pixels as well as radiometric overexposure effects that persisted despite atmospheric correction.

## 5 Conclusions

The high spatial and temporal resolution of the PlanetScope SuperDove cube satellite constellations reveals potential for agricultural monitoring applications including the retrieval of functional crop traits as the LAI. The potential is especially promising for small-scale agricultural structures like those found in Switzerland but also in other parts of the world. Cube satellites can thus complement established platforms such as Sentinel-2 for crop monitoring applications. However, challenges regarding radiometry, spatial co-registration and quality check should be addressed to enable operational use.

## 6 References

- DANNER, M., BERGER, K., WOCHER, M., MAUSER, W. & HANK, T., 2021: Efficient RTM-based training of machine learning regression algorithms to quantify biophysical & biochemical traits of agricultural crops. *ISPRS Journal of Photogrammetry and Remote Sensing*, **173**, 278-296, <https://doi.org/10.1016/j.isprsjprs.2021.01.017>.
- GRAF, L.V., PERICH, G. & AASEN, H., 2022: EOdal: An open-source Python package for large-scale agroecological research using Earth Observation and gridded environmental data. *Computers and Electronics in Agriculture*, **203**, 107487, <https://doi.org/10.1016/j.compag.2022.107487>.
- JACQUEMOUD, S., VERHOEF, W., BARET, F., BACOUR, C., ZARCO-TEJADA, P.J., ASNER, G.P., FRANÇOIS, C. & USTIN, S.L., 2009: PROSPECT+SAIL models: A review of use for vegetation characterization. *Remote Sensing of Environment, Imaging Spectroscopy Special Issue*, **113**, S56-S66, <https://doi.org/10.1016/j.rse.2008.01.026>.
- SCHEFFLER, D., HOLLSTEIN, A., DIEDRICH, H., SEGL, K. & HOSTERT, P., 2017: AROSICS: An Automated and Robust Open-Source Image Co-Registration Software for Multi-Sensor Satellite Data. *Remote Sensing*, **9**, 676, <https://doi.org/10.3390/rs9070676>.
- TU, Y.-H., JOHANSEN, K., ARAGON, B., EL HAJJ, M.M. & MCCABE, M.F., 2022: The radiometric accuracy of the 8-band multi-spectral surface reflectance from the planet SuperDove constellation. *International Journal of Applied Earth Observation and Geoinformation*, **114**, 103035, <https://doi.org/10.1016/j.jag.2022.103035>.
- WEISS, M., JACOB, F. & DUVEILLER, G., 2020: Remote sensing for agricultural applications: A meta-review. *Remote Sensing of Environment*, **236**, 111402, <https://doi.org/10.1016/j.rse.2019.111402>.

# Flying Parameters Experiment of a UAV LiDAR over a Winter Wheat Field-trial

CHRISTOPH HÜTT<sup>1</sup>

*Abstract: The study aimed to study how flying parameters of Unmanned Aerial Vehicle (UAV) Light Detection and Ranging (LiDAR) influences the retrieval of winter wheat plant height. Results showed that flight speed and altitude, which impact point density, affected how well the UAV LiDAR point clouds could be related to manual height measurements. Despite lower correlation at higher speeds and altitudes, the study found that plant height could still be accurately measured with increased efficiency at higher speeds and altitudes. A non-nadir look angle was found to have a higher correlation than a nadir look angle. However, further research is needed to understand better the relationship between UAV LiDAR flight parameters with plant height and other crop traits such as biomass and nitrogen uptake.*

## 1 Introduction

Integrating miniaturized laser scanning (LiDAR) systems on unmanned aerial vehicles (UAV), also known as UAV LiDAR systems, has opened up new possibilities for monitoring in various fields, including agriculture (e.g., SHENDRYK et al. 2020). Effective monitoring is crucial in agriculture due to global food shortages and the impacts of climate change (WEISS et al. 2020). However, the use of UAV LiDAR systems in agriculture is limited, partly due to a lack of understanding of flight parameters and their effect on the quality of the derived geospatial information.

In UAV LiDAR remote sensing, there is a trade-off between resolution and coverage. High-resolution, e.g., a dense point cloud, is limited in coverage, while lower-resolution data allows for more extensive coverage but may degrade information quality. Improving flight characteristics knowledge could enable improved planning for optimizing the monitoring of larger areas using UAV LiDAR remote sensing for agricultural monitoring.

Therefore, I conducted a UAV LiDAR experiment over a winter wheat field trial in Germany to study the relationship between flight parameters and plant height, a critical factor in precision agriculture applications.

## 2 Study Site and Field measurements

### 2.1 Study Site

The study took place at the Campus Klein-Altendorf (CKA), located about 20 km from Bonn. The study site was a winter wheat field trial managed with 120 plots, each 7m × 1.5m, arranged in 5 rows with three different nitrogen treatments (0, 120, 240 kg/ha) and six different winter wheat

---

<sup>1</sup> Universität zu Köln, Geographisches Institut, AG GIS & Fernerkundung, Albertus-Magnus-Platz, D-50923 Köln, E-Mail: christoph.huett@uni-koeln.de

varieties. The study site is equivalent to the studies done by JENAL et al. 2021 and HÜTT et al. 2022, where the field trial setup is described in more detail.

## 2.2 Manual Height Measurement of the Winter Wheat Plants

Figure 1 (left) shows where winter wheat plants were manually measured. The positions of the height measurements of winter wheat plants were measured using an RTK GPS Topcon GR-5 system in a Base Rover configuration. It provides sub-5 cm accuracy in the X, Y, and Z positions (TOPCON, 2023), and I used the same base station to correct the UAV LiDAR trajectory (section 2.4). A round plastic disc with a diameter of 30 cm was attached to the pole of the RTK GPS device. This plate was adjustable vertically along the pole and was positioned as high as the two highest wheat plants at each measurement location. The distance from the plate to a laser distance measurement device (Leica Disto) was then measured by this device and recorded using a smartphone. Subtracting the laser distance device's height from the measured distance between the plastic plate and the distance measurer allowed calculating the wheat plant heights. A photo of the device is shown in Fig.2 (right).

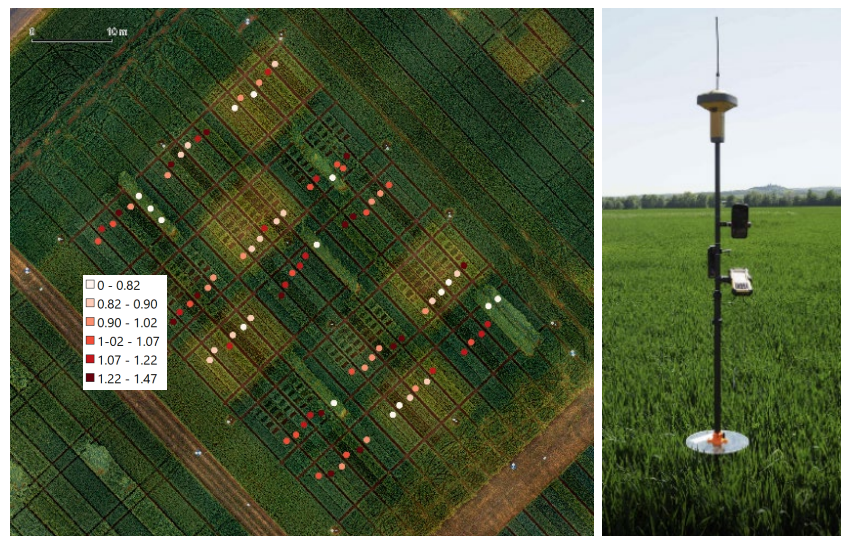


Fig. 1: Left: RGB orthoimage acquired with a DJI Phantom 4 UAV of the study site with the locations of the manual plant height measurements. Right: RTK GPS Laser Distance Winter Wheat Plant Height Measurer

## 2.3 UAV LiDAR Campaign

On June 23, 2021, the UAV LiDAR experiment was conducted with a Riegl Mini-VUX-1 mounted on a DJI Matrice 600 pro. The 360° rotating mirror of the LiDAR scanner reflects the beams around the UAV, perpendicular to its flight direction, providing up to 100,000 measurements per second (RIEGL, 2023). As a result, the area below the UAV is scanned in a line scanner pattern. UgCS was used to preplan and automate the flight. The aim was to achieve uniform point spacing in both range and azimuth directions by adjusting the flight height (to control range distance between points) and speed (to control azimuth distance between points). At first, the UAV was flown lower (Fig 2, middle) and slower (Fig.2 left) to achieve a denser point cloud. Later, faster and higher overflights of the field were carried out.

Furthermore, the faster and higher the UAV was planned to fly, the more space was given, before and after the field, to allow the UAV to have a constant speed over the field trial (Fig.2 right). To also examine the effect of the look angle, the field was overflown directly above it and with an offset, so the field trial is viewed approximately with an off-nadir angle of 45°. Table 1 lists the flight characteristics, an estimation of point spacing, and potential area coverage per minute. The area coverage calculation assumes that only points at an angle up to 45° off-nadir are used and assumes a straight flight path without turns, thus overestimating the actual coverage.

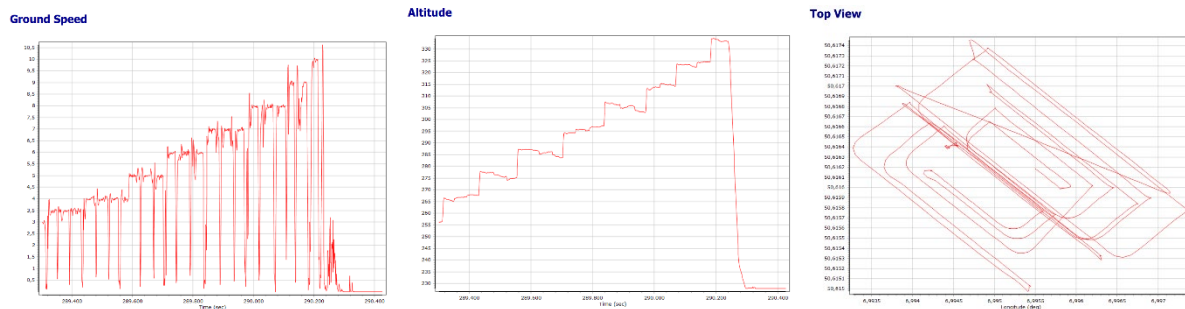


Fig. 2: Plots of the trajectory: left: increasing ground speed, middle: increasing flight height, right: Top View of the trajectory

Tab. 1: Flight Plan Characteristics, \* at 110 m, there was no off-nadir flight

Height above ground (m)	Ground Speed (m/s)	approx. Dist. between Points (m)	Swath Width (+/- 40°) (m)	potential Area Coverage/Minute (ha)
30	2	0.07	50.3	0.6
40	3	0.08	67.1	1.2
50	4	0.12	83.9	2.0
60	5	0.14	100.7	3.0
70	6	0.16	117.5	4.2
80	7	0.19	134.3	5.6
90	8	0.21	151.0	7.2
100	9	0.23	167.8	9.1
110*	10	0.26	184.6	11.1

## 2.4 Generating and processing UAV LiDAR point clouds

The Applanix IMU APX-15 captured the trajectory and orientation of the UAV LiDAR system, and GPS correction data was obtained via the Topcon GNSS GR-5 standing next to the field. During post-processing, these datasets were integrated to approximate the system's location during the flight with high precision (0.02 - 0.05m) (APPLANIX 2023). Subsequently, a preliminary version of the LiDAR point cloud was generated in RiProcess based on the corrected trajectory. The accuracy of the point cloud was then enhanced further by incorporating ground targets into the analysis in RiPrecision. Finally, each overflight was separated and subjected to individual analysis.



## 2.5 Processing of the UAV LiDAR point clouds and Statistical analysis

The analysis of the individual point cloud strips was conducted in LASTools as follows:

1. Outlier points were removed if there were fewer than 4 points per square meter.
2. The height was normalized using a ground model from the same year (February 3) to obtain the relative height above ground instead of absolute heights.
3. A 15 cm buffer was created in a GIS environment around the manual height measurements, which had the same extent as the manual measurement plate.
4. The maximum height of points was calculated using the LAScanopy function of LASTools for each measurement area, with all points falling into the buffer.

As a final step, linear regressions were established between the maximum LiDAR height and the manually measured plant height for each overflight. Fig. 3 shows three normalized points from different height and speed combinations.

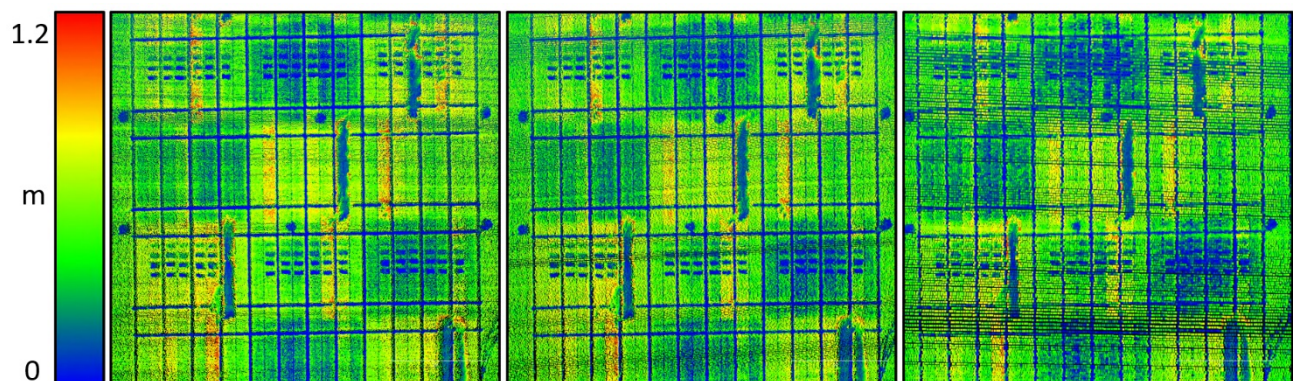


Fig. 3: Normalized point clouds from 3 different heights (all nadir), coloured by height above ground: left: 30m - 2m/s, middle: 60m – 5 m/s, right: 100 m – 9 m/s right

## 3 Results

The results show that, in general, higher flight heights with higher flight speeds and hence lower point densities result in a lower correlation with the manual height measurements (Tab.2). Interestingly, the average  $R^2$  from the off-nadir data is 0.035 higher than the  $R^2$  using data from the nadir overflights. The decrease of  $R^2$  from 30 m height and 2 m/s to 100 m and 9 m/s is about 0.05 on average, and the same for off-nadir and nadir look (Fig. 4). The highest correlation between manually measured plant height and maximum LiDAR height was obtained for the 50 m flight at 4m/s with a non-nadir look angle. Noteworthy are the intercepts of the linear regressions. In the nadir case, their average is 0.52 m, with an increasing tendency towards higher altitudes. For the off-nadir scenario, it is 0.4 m on average, without a clear trend.

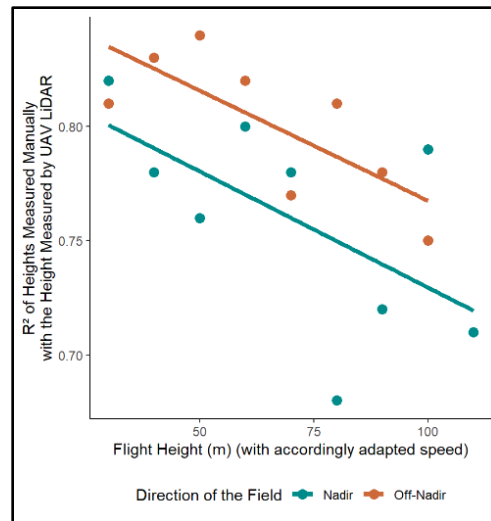


Fig. 4: The  $R^2$  values for 17 combinations of UAV flight height and speed for both nadir and off-nadir viewing angles

Tab. 2: Correlation Coefficients of the maximum point height above ground of the different height and speed combinations and the manual measurements

Flight height	Nadir		45° Off-Nadir	
	$R^2$	Linear Regression	$R^2$	Linear Regression
30	0.82	$0.46 + 0.91x$	0.81	$0.42 + 0.90x$
40	0.77	$0.50 + 0.86x$	0.83	$0.41 + 0.90x$
50	0.78	$0.49 + 0.88x$	0.83	$0.38 + 0.94x$
60	0.79	$0.50 + 0.89x$	0.82	$0.41 + 0.91x$
70	0.78	$0.54 + 0.88x$	0.77	$0.38 + 0.94x$
80	0.69	$0.59 + 0.80x$	0.81	$0.38 + 0.99x$
90	0.73	$0.56 + 0.86x$	0.78	$0.44 + 0.95x$
100	0.79	$0.56 + 0.89x$	0.75	$0.43 + 0.88x$
110	0.71	$0.61 + 0.81x$		

## 4 Discussion

The study showed that it is possible to measure winter wheat plant height using UAV LiDAR accurately. This conclusion agrees with previous studies on the topic (TEN HARKEL et al. 2019; BATES et al. 2021; HÜTT et al. 2022). However, when UAV LiDAR is used at higher flying speeds and altitudes, the point density of the point clouds decreases, negatively impacting the correlation with manual measurements. Despite this, even with a coarser point cloud obtained from a 100 m altitude at 9m/s flying speed, the results are still considered acceptable, with  $R^2$  values of 0.79 for nadir measurements and 0.75 for off-nadir measurements.

The difference in correlation between nadir and off-nadir measurements is noteworthy and could be because, in a nadir look, it is more likely that the highest plants were missed, as winter wheat plants grow vertically. In this study, the manual measurements focused on the highest plants in a

circle with a 30 cm diameter. Further investigation into how different UAV LiDAR metrics (HÜTT et al. 2022) respond to the difference in look angle and point density would be valuable.

Additionally, repeating the experiment at different stages in the growing period would be interesting. Also, different UAV LiDAR systems, such as Riegl's VUX, which has a higher sampling rate and provides the full waveform of each LiDAR return, would be a great addition. Such research would allow for a more detailed and comprehensive analysis of the data and a deeper understanding of the relationship between UAV LiDAR and winter wheat plant height.

## 5 Conclusion and Outlook

The study shows that it is generally possible to use UAV LiDAR to monitor winter wheat plant heights with high accuracy and precision. The results indicate that flight speed and altitude, which influence the point density, negatively impacted the accuracy of the correlation with manual height measurements. However, if maximizing accuracy is not a primary concern, plant height can still be accurately determined at higher speeds and altitudes, resulting in increased efficiency. Further research is needed to consolidate the study's findings and how the different flight parameters influence the estimation of crop traits such as plant height, biomass, or nitrogen uptake. A refinement of future crop trait retrieval algorithms could be in considering the angle under which the agricultural area is viewed.

## 6 References

- APPLANIX, 2023: APX-15 UAV Datasheet. [https://www.applanix.com/downloads/products/specs/APX15\\_UAV.pdf](https://www.applanix.com/downloads/products/specs/APX15_UAV.pdf), last access 03.02.2023.
- BATES, J. S., MONTZKA, C., SCHMIDT, M. & JONARD, F., 2021: Estimating Canopy Density Parameters Time-Series for Winter Wheat Using UAS Mounted LiDAR. *Remote Sensing*, **13**(4), 710, <https://doi.org/10.3390/rs13040710>.
- HÜTT, C., BOLTEN, A., HÜGING, H. & BARETH, G., 2022: UAV LiDAR Metrics for Monitoring Crop Height, Biomass and Nitrogen Uptake: A Case Study on a Winter Wheat Field Trial. *PFG—Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 1-12, <https://doi.org/10.1007/s41064-022-00228-6>.
- JENAL, A., HÜGING, H., AHRENDT, H. E., BOLTEN, A., BONGARTZ, J. & BARETH, G., 2021: Investigating the Potential of a Newly Developed UAV-Mounted VNIR/SWIR Imaging System for Monitoring Crop Traits—A Case Study for Winter Wheat. *Remote Sensing*, **13**(9), 1697, <https://doi.org/10.3390/rs13091697>.
- RIEGL, 2023: Miniaturized LiDAR sensor for unmanned laser scanning RIEGL miniVUX-1 UAV ®. <http://www.riegl.com/products/unmanned-scanning/riegl-minivux-1uav/>, last access 05.02.2023.
- SHENDRYK, Y., SOFONIA, J., GARRARD, R., RIST, Y., SKOCAJ, D. & THORBURN, P., 2020: Fine-Scale Prediction of Biomass and Leaf Nitrogen Content in Sugarcane Using UAV LiDAR and Multispectral Imaging. *International Journal of Applied Earth Observation and Geoinformation*, **92**, 102177, <https://doi.org/10.1016/j.jag.2020.102177>.

TEN HARKEL, J., BARTHOLOMEUS, H. & KOOISTRA, L., 2019: Biomass and Crop Height Estimation of Different Crops Using UAV-Based LiDAR. *Remote Sensing*, **12**(1), 17, <https://doi.org/10.3390/rs12010017>.

TOPCON, 2023: GR-5 Advanced GNSS Receiver. [https://www.topconpositioning.com/sites/default/files/product\\_files/gr-5\\_broch\\_7010\\_2071\\_revj\\_sm.pdf](https://www.topconpositioning.com/sites/default/files/product_files/gr-5_broch_7010_2071_revj_sm.pdf), last access 5.2.2023.

WEISS, M., JACOB, F. & DUVEILLER, G., 2020: Remote Sensing for Agricultural Applications: A Meta-Review. *Remote Sensing of Environment*, **236**, 111402, <https://doi.org/10.1016/j.rse.2019.111402>.

# Instanzsegmentierung von stehendem Totholz mittels Mask R-CNN und CIR-Bildern

SEBASTIAN BRIECHLE<sup>1</sup>

*Zusammenfassung: Die Kartierung von stehendem Totholz auf Einzelbaumebene wird – u.a. aufgrund der sich ändernden klimatischen Bedingungen – zu einem immer bedeutenderen Aspekt im Zuge der Waldinventur. Jedoch fehlen bisher leistungsstarke Methoden, die auf Basis von flächenhaft zur Verfügung stehenden Fernerkundungsdaten präzise Ergebnisse auf großen Waldflächen liefern können. In dieser Arbeit wird gezeigt, dass die Deep Learning-basierte Instanzsegmentierung auf diese Problemstellung angepasst werden kann. Hierbei konnte auf Basis von 20 cm CIR-Orthophotos aus dem Nationalpark Bayerischer Wald (530 Bäume/ha) das neuronale Netz Mask R-CNN trainiert werden, das in der Lage ist, einzelne stehende tote Bäume zu detektieren und zudem die Umrisse der Baumkronen zu liefern. Des Weiteren wurde die Generalisierungsfähigkeit des trainierten Modelles anhand von Testdaten mit unterschiedlich starker Ähnlichkeit zum Trainingsdatensatz evaluiert. Generell betrug die Inferenzzeit für 1 ha große Bildkacheln 50 ms (5 s pro km<sup>2</sup>) und untermauert das enorme Potenzial dieser Methode für die großflächige Totholzkartierung in der Praxis.*

## 1 Einleitung

### 1.1 Motivation

Weltweit leiden die Wälder unter sich ändernden klimatischen Bedingungen, starken Winden und einem zunehmenden Schädlingsbefall. Vor diesem Hintergrund hat die flächendeckende Kartierung von abgestorbenen Bäumen mit Hilfe von Fernerkundungsdaten in den letzten Jahren immer mehr an Bedeutung gewonnen. In der Praxis haben großflächige Methoden, die vollautomatisch und zuverlässig Ergebnisse auf der Skala einzelner Bäume liefern können, ein großes Potenzial. Einerseits ist eine genaue quantitative Bewertung der ökologischen Situation von großem Interesse. Andererseits ist es für die von massivem Waldsterben betroffenen Waldbewirtschaftenden und deren Versicherungen wichtig, den wirtschaftlichen Schaden schnell, möglichst genau und vollautomatisch abzuschätzen. Noch fehlt es jedoch an robusten, zuverlässigen und zugleich genauen Methoden zur Totholzkartierung – insbesondere in natürlichen und damit komplexen Wäldern.

### 1.2 Totholzkartierung

Die Literaturrecherche zur Totholzkartierung mit Fernerkundungsdaten zeigt, dass die meisten Ansätze auf Baumebene entweder auf Lidardaten oder Multi-(MS)/Hyperspektralbildern, oder einer Kombination aus beiden, beruhen. Lidar-basierte Methoden zur Totholzkartierung sind in der Regel zweistufig. In einem ersten Schritt erfolgt die Abgrenzung von Einzelbäumen – entweder mittels *canopy height model* (CHM) (PYYSALO & HYYPPÄ 2002) oder punktwolken-basiert (REITBERGER et al. 2009). In einem zweiten Schritt werden die Einzelbäume als „tot“ oder „lebendig“

---

<sup>1</sup> Hochschule München, Fakultät für Geoinformation, Labor für Photogrammetrie und Fernerkundung, Karlstraße 6, D-80333 München, E-Mail: sebastian.briechle@hm.edu

klassifiziert. Hier werden häufig manuell generierte Merkmale und gängige Methoden des maschinellen Lernens (z.B. *random forest*, *support vector machine*) verwendet (KAMINSKA et al. 2018; KRZYSZEK et al. 2020). In letzter Zeit haben tiefe neuronale Netze (DNNs) diese Methoden weitgehend ersetzt (HAMRAZ et al. 2019; BRIECHLE et al. 2020), da sie in der Lage sind, charakteristische Merkmale automatisch aus den Rohdaten zu extrahieren. Im Gegensatz zu Lidar-basierten Methoden machen sich rein bildbasierte Ansätze die Tatsache zunutze, dass sich Totholz insbesondere in *color-infrared* (CIR)-Bildern deutlich von der lebenden Vegetation abhebt. CIR-Orthophotos werden flächendeckend durch die Landesämter bereitgestellt, sie sind vergleichsweise günstig und haben zudem eine hohe geometrische Auflösung (meist 20 cm). Jüngste Veröffentlichungen haben gezeigt, dass *convolutional neural networks* (CNNs) aus dem Computer-Vision-Bereich erfolgreich für Orthophotos aus dem Forstbereich verwendet werden können. SYLVAIN et al. (2019) zeigten beispielsweise das Potenzial von CNNs zur Kartierung des Gesundheitszustandes von Bäumen („lebend“ bzw. „tot“). Unter Verwendung von 20 cm-Luftbildern erzielten die Autoren eine Testgenauigkeit auf Pixelebene von bis zu 94%.

### 1.3 Instanzsegmentierung

Neben DNNs für die Bildklassifizierung, semantische Segmentierung und Objekterkennung sind Modelle, die eine Instanzsegmentierung durchführen, von zunehmendem Interesse. Der größte Vorteil dieser Modelle für forstwirtschaftliche Anwendungen liegt auf der Hand: Sie ermöglichen die Abgrenzung und Klassifizierung einzelner Baumkronen in einer durchgängigen Pipeline. In letzter Zeit wurden einige Arbeiten veröffentlicht, die Mask R-CNN (HE et al. 2017) zur Detektion und Klassifizierung von Einzelbäumen verwendeten. BRAGA et al. (2020) nutzten Mask R-CNN zur Erkennung von Baumkronen in sehr hochauflösenden Satellitenbildern von tropischen Wäldern ( $F_1$ -Score = 0,86). Im selben Jahr setzten CHADWICK et al. (2020) Mask R-CNN auf UAV-basierten RGB-Bildern (GSD = 3 cm) ein ( $F_1$ -Score = 0,91). HAO et al. (2021) trainierten Mask R-CNN zur Erkennung von Baumkronen in hochauflösenden UAV-basierten MS-Bildern einer Plantage in China. Ein Modell mit einer Kombination aus *normalized difference vegetation index* (NDVI) und CHM erzielte die besten Ergebnisse ( $F_1$ -Score = 0,85).

Im Gegensatz zur Klassifizierung von Baumarten wurde über die DL-basierte Instanzsegmentierung zur Totholzkartierung bisher nur sehr wenig berichtet. Daher wird hier nur auf eine Veröffentlichung verwiesen: POLEWSKI et al. (2021) stellten ein zweistufiges Verfahren zur Instanzsegmentierung einzelner umgestürzter Bäume vor. Auf Basis einer semantischen Segmentierung, die mittels U-Net (RONNEBERGER et al. 2015) erstellt wurde, wird ein aktives Konturverfolgungsverfahren genutzt. Unter Verwendung von hochauflösenden CIR-Orthophotos erreichten die Autoren einen  $F_1$ -Score von bis zu 0,87 für Testflächen im Nationalpark Bayerischer Wald (NPBW).

### 1.4 Kernidee

Nach bestem Wissen des Autors wurde Mask R-CNN noch nie zur Kartierung einzelner stehender toter Bäume in einem natürlichen und damit anspruchsvollen Waldgebiet verwendet. Daher besteht der Hauptbeitrag dieser Arbeit darin, das Potenzial der Anwendung von Mask R-CNN auf diesen Praxisfall hinsichtlich Genauigkeit und Flächenleistung zu untersuchen.

## 2 Materialien und Methoden

### 2.1 Untersuchungsgebiet

Das Untersuchungsgebiet für die folgenden Experimente befindet sich im NPBW nahe der deutsch-tschechischen Grenze und umfasst etwa 530 Bäume pro ha. Der NPBW wurde 1970 eingerichtet und die dominierenden Baumarten sind Fichten (*Picea albies*), Rotbuchen (*Fagus sylvatica*), Weißtannen (*Abies alba*) und Lärchen (*Larix*). Darüber hinaus haben natürliche Belastungen – schwere Stürme, Borkenkäferbefall und lange Trockenperioden – zu großen Mengen an stehendem Totholz geführt.

### 2.2 Datensatz und Vorverarbeitung

Im Juni 2016 wurden CIR-Luftbilder mit einer Leica DMC III Kamera (Brennweite 92 mm) aufgenommen (LATIFI et al. 2021). Unter Verwendung eines hochaufgelösten Lidar-basierten digitalen Oberflächenmodelles (DOM) wurden anschließend 20 cm CIR-Orthophotos erstellt. In einem nächsten Schritt wurde das Orthomosaik in Kacheln der Größe  $512 \times 512$  Pixel aufgeteilt. Damit entspricht die Bildgröße in etwa 1 ha.

Anschließend wurden Referenzdaten auf der Grundlage einer visuellen Interpretation manuell annotiert. In der vorliegenden Arbeit wurden abgestorbene Bäume mit Hilfe des Open-Source-Tools LabelMe (RUSSELL et al. 2007) als Kreise unterschiedlicher Größe markiert. Schließlich wurden die Bildkacheln zufällig in Trainings-, Validierungs- und Testdatensätze sortiert (siehe Tabelle 1).

Tab. 1: Übersicht über die Größe des Trainings-, Validierungs- und Testdatensatzes.

Datensatz	Anzahl der Bilder	Anzahl der Instanzen	Fläche [km <sup>2</sup> ]
train	131 (49%)	681 (43%)	1,37
validation	52 (19%)	486 (31%)	0,55
test	86 (32%)	407 (26%)	0,90
$\Sigma$	269	1574	2,82

### 2.3 Instanzsegmentierung mit Mask R-CNN

Die eigentliche Totholzkartierung wurde mit Detectron2 (WU et al. 2019), einer beliebten PyTorch-basierten Bibliothek durchgeführt. Detectron2 ermöglicht die Einbindung von State-of-the-Art-Technologien in den Workflow und umfasst Modelle wie Faster R-CNN, Mask R-CNN, RetinaNet und DensePose. Um die Detectron2-Pipeline nutzen zu können, muss in einem Vorverarbeitungsschritt ein benutzerdefinierter COCO-Datensatz aus den oben genannten Bildkacheln und Beschriftungen erstellt werden. Dies kann mit Hilfe der Funktion `register_coco_instances()` geschehen.

Als Modell wurde Mask R-CNN mit einem ResNet-50-Backbone zur Merkmalsextraktion und einem Feature Pyramid Network gewählt (`mask_rcnn_R_50_FPN_3x.yaml`), das *bounding boxes* und detaillierte Umrisse einzelner Baumkronen liefert. Tabelle 2 zeigt einige der wichtigsten Hyperparameter und ihre gewählten Werte. Der Code wurde als Jupyter-Notebook auf Google Colab ([colab.research.google.com](https://colab.research.google.com)) unter Verwendung von CUDA 10.1 und PyTorch 1.7, sowie einer NVIDIA® Tesla T4 GPU (2.560 CUDA® Cores, 16 GB GPU-Speicher) ausgeführt.

Tab. 2: Übersicht über die wichtigsten Parameter von Mask R-CNN in Detectron2.

	Parameter	Bedeutung	gesetzte Werte
model	<i>backbone.freeze_at</i>	einzufrierende ResNet-50-Blöcke (je größer, desto weniger Parameter werden neu trainiert)	1, 2, 3
	<i>anchor_generator.sizes</i>	Größe der Anker in [pix]	{8 16 32 64 128}
	<i>anchor_generator.aspect_ratios</i>	Seitenverhältnisse der Anker	{0,75 1,00 1,33}
solver	<i>ims_per_batch</i>	Anzahl der Bilder pro Stapel	8
	<i>base_lr</i>	initiale Lernrate	0,001
eval	<i>eval_period</i>	Intervall für die Validierung	50

Eines der wichtigsten Ziele beim Training eines neuronalen Netzes ist es, eine Überanpassung (*overfitting*) an die Trainingsdaten zu vermeiden. In der Praxis bedeutet dies, den Trainingsprozess im richtigen Moment zu stoppen. Andernfalls wird das Modell zu sehr für Vorhersagen auf den Trainingsdaten optimiert und verliert seine Generalisierungsfähigkeit auf neuen Daten. Daher wurde ein Abbruchkriterium (*early stopping*) eingeführt, das den *validation loss* überwacht (*patience* = 10), um sicherzustellen, dass das beste Modell gespeichert wird. Um das trainierte Modell so robust wie möglich gegenüber geometrischen und radiometrischen Veränderungen in den Eingangsbildern zu machen und damit die Übertragbarkeit auf neue Datensätze zu erhöhen, wurden die Daten während des Trainings künstlich angereichert (*data augmentation*). Einerseits wurden die Helligkeit, der Kontrast und die Sättigung der Bilder zufällig verändert, um die radiometrischen Eigenschaften zu variieren. Andererseits wurden die Bilder zufälligen geometrischen Transformationen unterzogen (horizontales und vertikales Spiegeln, Drehen und Zuschneiden).

Unter Verwendung des trainierten Modelles wurden Vorhersagen auf unabhängigen Testdatensätzen getroffen und mit den Referenzdaten verglichen. Zur quantitativen Bewertung wurde hierbei der sogenannte AP50-Wert ermittelt. AP50 ist die durchschnittliche Präzision bei einem *intersection over union* (IoU)-Schwellenwert von 50% – eine weit verbreitete Metrik im Bereich der Objekterkennung. Die Metrik wird sowohl für die Erkennung von einzelnen toten Baumobjekten (*AP50\_bbox*), als auch die Trennung von Hintergrund und Objekt (*AP50\_seg*) berechnet. Zur qualitativen Analyse wurden die Vorhersagen auf zufällig ausgewählten Testbildern mit den entsprechenden Eingabebildern überlagert und mit den Referenzdaten verglichen.

## 2.4 Generalisierung

Eine der größten Schwierigkeiten im Zusammenhang mit neuronalen Netzen ist es, einen hohen Grad an Generalisierung und damit eine hohe Übertragbarkeit auf neue Datensätze zu erreichen. Um das trainierte Netz hinsichtlich dieser Fähigkeit beurteilen zu können, wurden zwei weitere CIR-Testdatensätze aus dem NPBW generiert (siehe Tabelle 4). Diese weisen zwar dieselbe Vorverarbeitung und geometrische Auflösung auf wie der bereits angesprochene Trainingsdatensatz (20 cm), unterscheiden sich von diesem jedoch unterschiedlich stark in Bezug auf den Bildinhalt. Die Unterteilung in die Datensätze *test\_similar* und *test\_tough* erfolgte visuell mittels manueller Sortierung. Die anschließende quantitative und qualitative Evaluierung wurde identisch zur Evaluierung des Testdatensatzes durchgeführt (siehe Kapitel 2.3).



### 3 Ergebnisse und Diskussion

#### 3.1 Flächenleistung

Der Erfolg eines automatisierten Ansatzes in der Praxis hängt nicht nur von der Genauigkeit der Vorhersagen ab, sondern vor allem auch von der erreichbaren Flächenleistung. Unter Verwendung eines Jupyter-Notebooks auf Google Colab betrug die Trainingszeit pro Epoche 15 s und dementsprechend 72 min für ein komplettes Training (300 Epochen). Die Inferenzzeit – also die Zeit, die das trainierte Modell für die Vorhersage auf Bildkacheln der Größe von 1 ha benötigt – betrug 50 ms. Dies verdeutlicht die hohe Skalierbarkeit des Ansatzes für die großflächige Kartierung von Totholz: Eine Fläche von 1 km<sup>2</sup> kann in etwa 5 s kartiert werden, die gesamte Fläche des NPBW (248,5 km<sup>2</sup>) in weniger als einer halben Stunde (21 Minuten).

#### 3.2 Quantitative Auswertung

Mit Hilfe der oben vorgestellten Pipeline wurden verschiedene Experimente durchgeführt, um ein optimiertes Mask R-CNN-Modell zu erhalten. Als Auswahlkriterium diente der niedrigste Wert für den *validation loss*, was ein übliches Vorgehen ist. Auf den 86 unabhängigen Testbildern (mit 407 toten Bäumen und einer Fläche von 90 ha) erreichte das beste Modell einen *AP50\_bbox* Wert von 72,6 und einen *AP50\_seg* Wert von 73,0 (siehe Tabelle 3).

Generell ist festzustellen, dass die besten Ergebnisse mit einem Wert für *backbone.freeze\_at* von 2 erzielt wurden, was der Standardeinstellung in Detectron2 entspricht. Darüber hinaus konnten schrittweise Verbesserungen mit einer höheren Intensität der geometrischen (*geom*) und radiometrischen (*radiom*) Datenanreicherung erreicht werden. Der Zusatz *heavy* bedeutet zum einen, dass hier die Intervalle der Parameter für die Datenanreicherung größer als in den anderen Experimenten eingestellt wurden – außerdem wurden die Eingabebilder während des Trainings zufällig zugeschnitten.

Tab. 3: Ergebnis der Instanzsegmentierung auf dem Testdatensatz (86 Kacheln, 407 tote Bäume) für eine unterschiedlich hohe Anzahl an geschätzten Parametern im Zuge der Merkmalsextraktion und eine unterschiedlich starke Intensität der Datenanreicherung während des Trainings; bestes Ergebnis in fett.

<b>backbone.freeze_at</b>	<b>Datenanreicherung</b>	<b>AP50_bbox</b>	<b>AP50_seg</b>
1	geom	67,0	65,9
1	geom, radiom	68,4	69,4
2 (default)	geom	66,8	69,0
2 (default)	geom, radiom	68,6	71,5
2 (default)	geom_heavy, radiom	<b>72,6</b>	<b>73,0</b>
3	geom	61,1	59,1
3	geom, radiom	61,5	60,3

#### 3.3 Qualitative Auswertung

Zur qualitativen Analyse des besten Modelles zeigt Abbildung 1 beispielhafte Bildkacheln (links), Referenzdaten (Mitte) und Vorhersagen (rechts). Offensichtlich wurden die meisten abgestorbenen Bäume von Mask R-CNN erfolgreich erkannt und umrissen. Obwohl kreisförmige Masken für die Annotationen verwendet wurden, war das DNN in der Lage, die Umrisse der einzelnen Bäume zu lernen. Die visuelle Interpretation zeigt auch, dass die Größe der Baumkronen in den

meisten Fällen korrekt geschätzt werden konnte. Diese Tatsache eröffnet die Möglichkeit, in zukünftigen Experimenten Kronendurchmesser aus den Polygonflächen abzuschätzen.

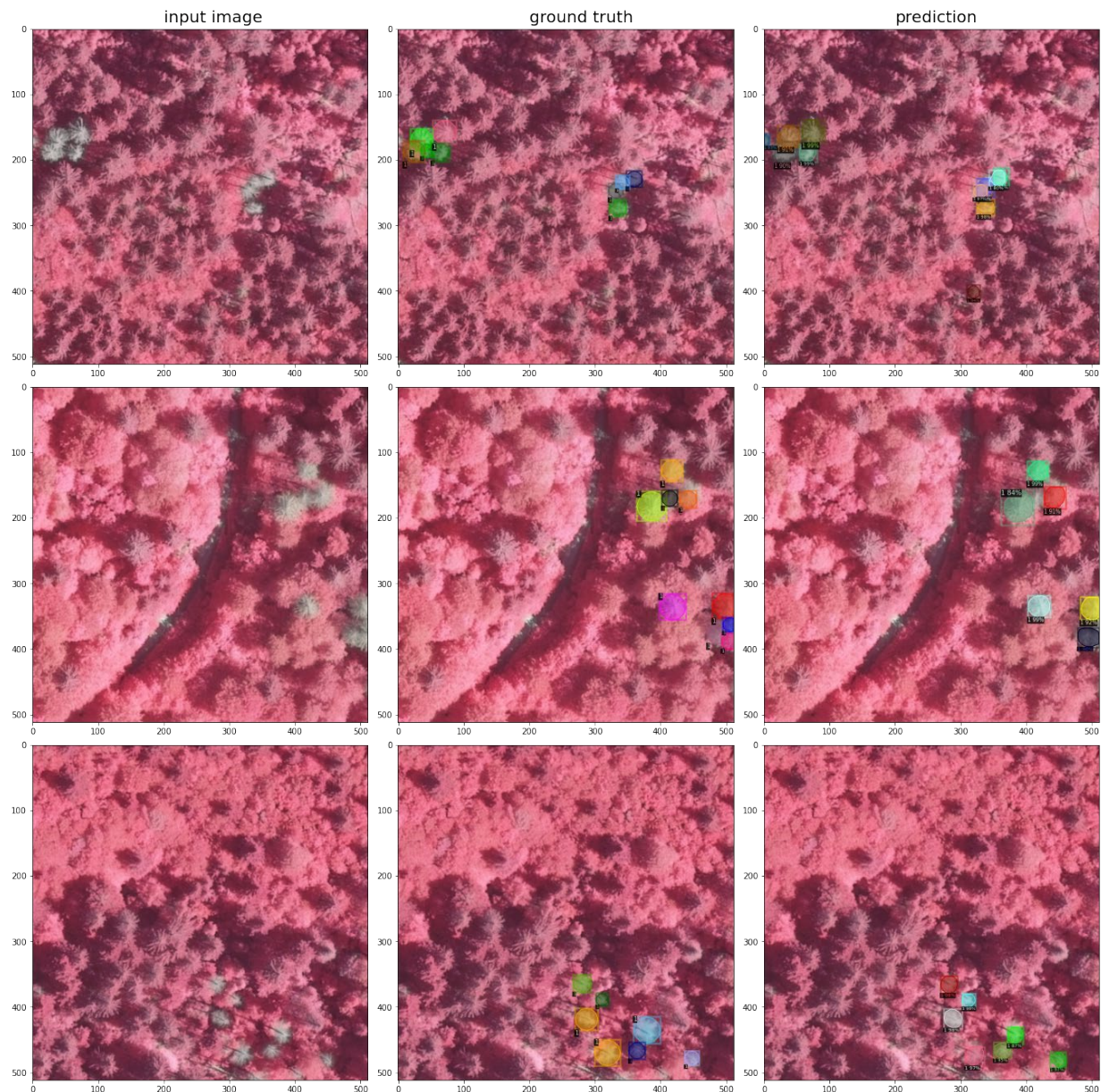


Abb. 1: Beispiele für eine erfolgreiche Totholzkartierung in verschiedenen Szenarien unter Verwendung von Mask R-CNN; Eingabebild (links), Referenzdaten (Mitte), Vorhersage (rechts).

Dennoch zeigen einige Beispiele von Fehlklassifikationen, dass der vorgestellte Ansatz noch Verbesserungspotenzial hat. In einzelnen Fällen produzierte das DNN falsch-positive Ergebnisse bei teilweise beschatteten Straßen, bei denen das CIR-Bild geometrische und radiometrische Strukturen aufweist, die denen eines toten Baumes sehr ähnlich sind (Abbildung 2, obere Reihe). Hier könnte die nachträgliche Filterung mittels Wegepolygonen bzw. die Integration der relativen Höheninformation (nDOM) Abhilfe schaffen.

Darüber hinaus traten vor allem an den Bildrändern vermeintliche falsch-positive Fehlklassifizierungen auf (Abbildung 2, untere Reihe). Hierbei handelt es sich um offensichtliche Fehler in den Referenzdaten – das DNN traf eigentlich eine korrekte Vorhersage. Eine Lösung für das Problem der abgeschnittenen Objekte an den Kachelrändern und der daraus resultierenden Fehler bei der Erkennung liegt auf der Hand: Bei der Erstellung der Eingangsdaten sollte eine ausreichend große Überlappung zwischen benachbarten Kacheln sichergestellt werden. Anschließend könnten die einzelnen abgeschnittenen Polygone aufgrund der geringeren Fläche im Vergleich zum komplett erfassten Einzelbaum in einem Nachbearbeitungsschritt gefiltert werden.

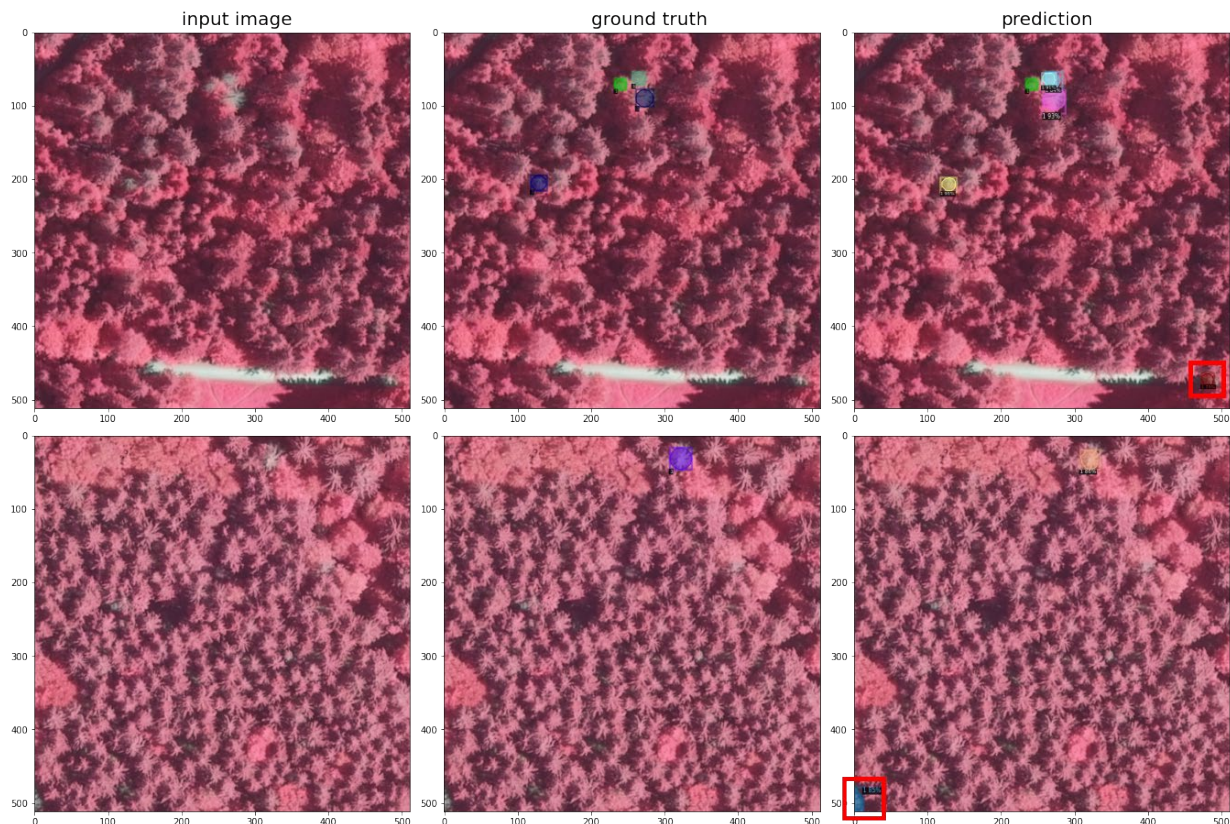


Abb. 2: Auftretende Fehlklassifizierungen aufgrund (i) einer teilweise schattierten Straße (obere Reihe) und (ii) eines fehlenden Labels in den Referenzdaten am Bildrand (untere Reihe); rotes Quadrat markiert Fehlklassifizierungen; Eingabebild (links), Referenzdaten (Mitte), Vorhersage (rechts).

### 3.4 Übertragbarkeit

Die Fähigkeit des trainierten Modelles zur Generalisierung – und damit zur Übertragbarkeit auf neue Datensätze – wurde anhand von zwei weiteren Testdatensätzen aus dem NPBW überprüft, deren Bildinhalte eine unterschiedlich starke Ähnlichkeit mit denen des Trainingsdatensatzes aufweisen (siehe Kapitel 2.4). Dabei wurde das Modell ausgewählt, das auf den ursprünglichen Testdaten die besten Ergebnisse lieferte ( $AP50_{bbox} = 72,6$ ;  $AP50_{segm} = 73,0$ ; siehe Tabelle 3). Wie zu erwarten war, zeigt die quantitative Auswertung ein unterschiedliches Bild für die beiden Datensätze *test\_similar* und *test\_tough*.

Auf Szenen mit einer Komplexität, die mit der des Trainingsdatensatzes vergleichbar ist (*test\_similar*), konnten zufriedenstellende Ergebnisse erzielt werden (siehe Tabelle 4 und Abbildung 3

oben). Hier waren die Metriken für die Detektion lediglich um 5,2 (*AP50\_bbox*) bzw. für die Maskierung um 3,6 (*AP50\_seg*) niedriger als auf dem 90 ha großen Testdatensatz.

Auf dem Testdatensatz mit von den Trainingsdaten deutlich abweichendem Bildinhalt (*test\_tough*; siehe Abbildung 3 unten) zeigt sich, dass das trainierte neuronale Netz Schwierigkeiten hat, korrekte Vorhersagen im Rahmen der Totholzerkennung zu treffen (siehe Tabelle 4). Die entsprechenden Metriken liegen hier ca. 16 bis 18 Punkte unter dem Ergebnis, das auf dem Datensatz *test\_similar* erzielt wurde. Die Ergebnisse zeigen also deutlich, dass in zukünftigen Experimenten die Diversität des Trainingsdatensatzes hinsichtlich der Bildinhalte erhöht werden sollte, um die Übertragbarkeit zu steigern.

Tab. 4: Ergebnisse hinsichtlich der Fähigkeit zur Generalisierung des trainierten Modelles auf Datensätze mit unterschiedlich stark von den Trainingsdaten abweichenden Bildinhalten.

Datensatz	Anzahl der Bilder	Anzahl der Instanzen	Fläche [ha]	AP50_bbox	AP50_seg
<i>test_similar</i>	16	142	17	67,4	69,4
<i>test_tough</i>	37	489	39	49,5	52,6

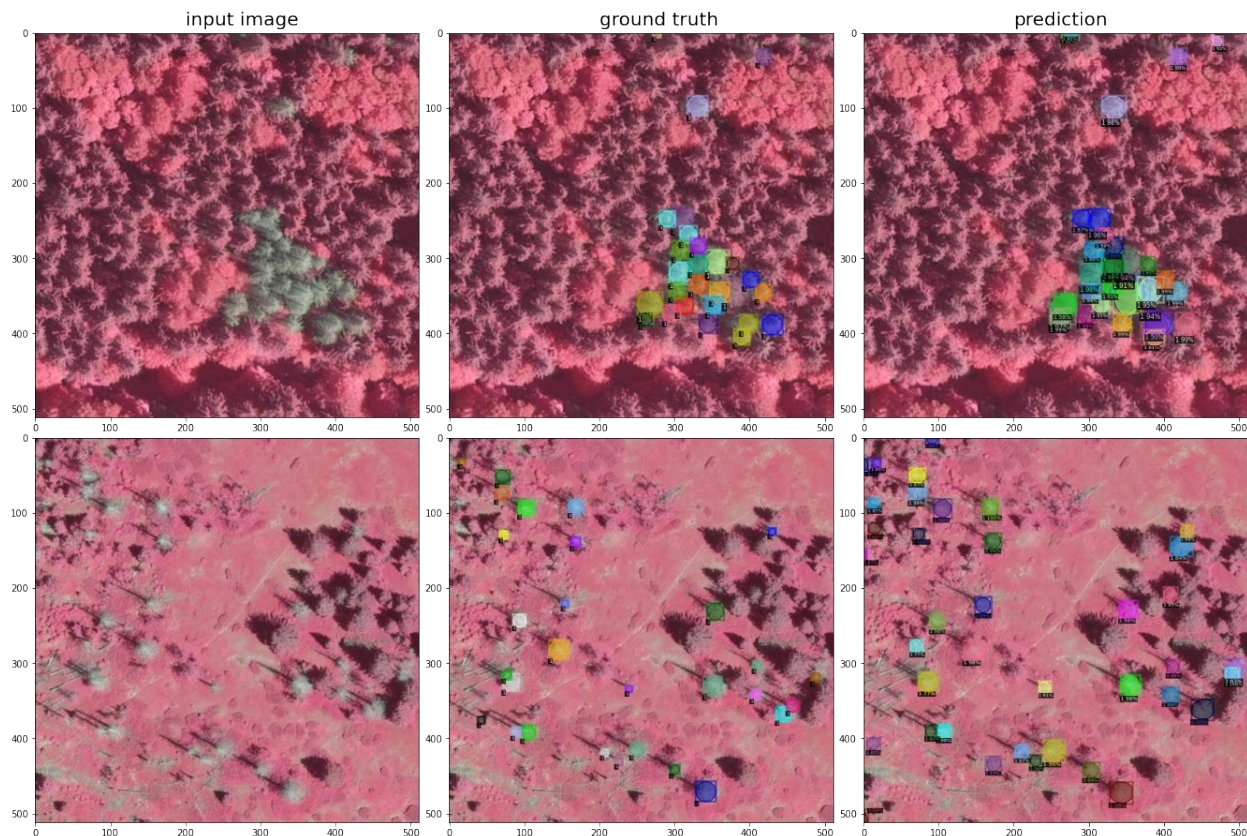


Abb. 3: Ergebnis der Instanzsegmentierung auf je einem repräsentativen Beispielbild der Datensätze *test\_similar* (1. Zeile) und *test\_tough* (2. Zeile); Eingabebild (links), Referenzdaten (Mitte), Vorhersage (rechts).

## 4 Fazit und Ausblick

In dieser Arbeit wurde Mask R-CNN – ein tiefes neuronales Netz zur Instanzsegmentierung – zur Totholzkartierung auf Einzelbaumebene verwendet. Die Wirksamkeit des vorgestellten Ansatzes wurde anhand eines 20 cm CIR-Orthomosaiks aus einem natürlichen Waldgebiet (530 Bäume/ha) demonstriert. Das beste Modell erreichte auf unabhängigen Testdaten einen *AP50\_bbox* Wert von 72,6 und einen *AP50\_segm* Wert von 73,0. Insgesamt bestätigen die Ergebnisse das große Potenzial für die Kartierung einzelner toter Bäume mit Mask R-CNN. Außerdem war die Leistung von Mask R-CNN in Detectron2 praxistauglich: Auf einer NVIDIA® Tesla T4 GPU (16 GB) betrug die Trainingszeit für den vorgestellten Datensatz 72 Minuten (300 Epochen) und die Inferenzzeit für 1 ha große Bildkacheln von 50 ms. Damit ist der vorgestellte Ansatz prinzipiell für die großflächige Kartierung von Totholz geeignet: Ein Gebiet von 1 km<sup>2</sup> kann in 5 s kartiert werden, der komplette NPBW (248,5 km<sup>2</sup>) in 21 min. Somit können flächendeckende und zuverlässige Karten für die Endnutzer erstellt werden, die für Anwendungen wie die automatisierte Waldinventur und Monitoringprojekte von großer Bedeutung sind. Für die Zukunft sind die folgenden Verbesserungen geplant:

- Generierung verbesserter, erweiterter und vielfältigerer Referenzdaten, indem man sich auf professionelle Felddaten und/oder Annotationen von Forstexpert:innen stützt;
- Integration von Wegedaten und/oder der relativen Höheninformation (nDOM), um die Anzahl der Fehlklassifizierungen im Postprocessing zu reduzieren und damit die Ergebnisse zu verbessern;
- Durchführung detaillierter Studien zur Übertragbarkeit auf weitere Untersuchungsgebiete (u.a. Frankenwald, Schwarzwald, Harz) und Eingangsdaten wie z. B. CIR-Orthophotos der Landesämter oder hochauflösende Satellitenbilder;

Grundsätzlich könnte die vorgestellte Methode auch auf die kombinierte Klassifizierung von Totholz und einzelnen Baumarten ausgeweitet werden. Allerdings sollte aufgrund der hohen Komplexität dieser Aufgabe eine höhere Anzahl von MS-Kanälen und ein Lidar-basiertes nDOM in die Pipeline aufgenommen werden.

## 5 Literaturverzeichnis

- BRAGA, G., PERIPATO, V., DALAGNOL, R., FERREIRA, M.P., TARABALKA, Y., ARAGÃO, L., DE CAMPOS VELHO, H.F., SHIGUEMORI, E.H. & WAGNER, F.H., 2020: Tree Crown Delineation Algorithm Based on a Convolutional Neural Network. *Remote Sensing*, **12**, 1288, <https://doi.org/10.3390/rs12081288>.
- BRIECHLE, S., KRZYSZEK, P. & VOSSELMAN, G., 2020: Classification of tree species and standing dead trees by fusing UAV-based lidar data and multispectral imagery in the 3D deep neural network PointNet++. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, **V-2-2020**, 203-210, <https://doi.org/10.5194/isprs-annals-V-2-2020-203-2020>.
- CHADWICK, A.J., GOODBODY, T., COOPS, N., HERVIEUX, A., BATER, C.W., MARTENS, L.A., WHITE, B. & RÖESER, D., 2020: Automatic Delineation and Height Measurement of Regenerating Conifer Crowns under Leaf-Off Conditions Using UAV Imagery. *Remote Sensing*, **12**, <https://doi.org/10.3390/rs12244104>.

- HAMRAZ, H., JACOBS, N., CONTRERAS, M. & CLARK, C., 2019: Deep learning for conifer/deciduous classification of airborne lidar 3D point clouds representing individual trees. *ISPRS Journal of Photogrammetry and Remote Sensing*, **158**, 219-230, <https://doi.org/10.1016/j.isprsjprs.2019.10.011>.
- HAO, Z., LIN, L., POST, C.J., MIKHAILOVA, E.A., LI, M., CHEN, Y., YU, K. & LIU, J., 2021: Automated tree-crown and height detection in a young forest plantation using mask region-based convolutional neural network (Mask R-CNN). *ISPRS Journal of Photogrammetry and Remote Sensing*, **178**, 112-123, <https://doi.org/10.1016/j.isprsjprs.2021.06.003>.
- HE, K., GKIOXARI, G., DOLLAR, P. & GIRSHICK, R., 2017: Mask R-CNN. In: *Proceedings of CVPR*, 2980-2988, <https://doi.org/10.1109/ICCV.2017.322>.
- KAMINSKA, A., LISIEWICZ, M., STERENCZAK, K., KRASZEWSKI, B. & SADKOWSKI, R., 2018: Species-related single dead tree detection using multi-temporal ALS data and CIR imagery. *Remote Sensing of the Environment*, **219**, 31-43, <https://doi.org/10.1016/j.rse.2018.10.005>.
- KRZYSZEK, P., SEREBRYANYK, A., SCHNÖRR, C., ČERVENKA, J. & HEURICH, M., 2020: Large-scale mapping of tree species and dead trees in Šumava National Park and Bavarian Forest National Park using lidar and multispectral imagery. *Remote Sensing*, **12**, <https://doi.org/10.3390/rs12040661>.
- LATIFI, H., HOLZWARTH, S., SKIDMORE, A., BRŮNA, J., ČERVENKA, J., DARVISHZADEH, R., HAIS, M., HEIDEN, U., HOMOLOVÁ, L., KRZYSZEK, P., SCHNEIDER, T., STARÝ, M., WANG, T., MÜLLER, J. & HEURICH, M., 2021: A laboratory for conceiving Essential Biodiversity Variables (EBVs)—The ‘Data pool initiative for the Bohemian Forest Ecosystem’. *Methods in ecology and evolution*, **12**(11), 2073-2083, <https://doi.org/10.1111/2041-210X.13695>.
- POLEWSKI, P., SHELTON, J., YAO, W. & HEURICH, M., 2021: Instance segmentation of fallen trees in aerial color infrared imagery using active multi-contour evolution with fully convolutional network-based intensity priors. *ISPRS Journal of Photogrammetry and Remote Sensing*, **178**, 297-313, <https://doi.org/10.1016/j.isprsjprs.2021.06.016>.
- PYYSALO, U. & HYYPPÄ, H., 2002: Reconstructing tree crowns from laser scanner data for feature extraction. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, **34**.
- REITBERGER, J., SCHNÖRR, C., KRZYSZEK, P. & STILLA, U., 2009: 3D segmentation of single trees exploiting full waveform lidar data. *ISPRS Journal of Photogrammetry and Remote Sensing*, **64**, 561-574, <https://doi.org/10.1016/j.isprsjprs.2009.04.002>.
- RONNEBERGER O., FISCHER P. & BROX T., 2015: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science*, **9351**, Springer, Cham, [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28).
- RUSSELL, B., TORRALBA, A., MURPHY, K. & FREEMAN, W.T., 2007: LabelMe: a database and web-based tool for image annotation. *International Journal of Computer Vision*, 2007.
- SYLVAIN, J.-D., DROLET, G. & BROWN, N., 2019: Mapping dead forest cover using a deep convolutional neural network and digital aerial photography. *ISPRS Journal of Photogrammetry and Remote Sensing*, **156**, 14-26, <https://doi.org/10.1016/j.isprsjprs.2019.07.010>.
- WU, Y., KIRILLOV, A., MASSA, F., LO, W.-Y. & GIRSHICK, R., 2019: Detectron2. <https://github.com/facebookresearch/detectron2>.

# Detection of Forest Fire Consequences on Satellite Images using a Neural Network

**VIKTORIIA HNATUSHENKO<sup>1,3</sup>, VOLODYMYR HNATUSHENKO<sup>2</sup>, VITA KASHTAN<sup>2</sup>  
& CHRISTIAN HEIPKE<sup>3</sup>**

*Abstract: The objective of this research is the detection of burnt forest areas from Sentinel-2 imagery. The proposed algorithm uses an approach based on convolutional neural networks (CNN). The functionality of the created system allows solving the task, starting from the moment of receiving the input data, image preprocessing and ending with the export of a hot-spot fire polygonal file describing the area that was burnt. These results are compared to methods based on the dNBR and a variant of BAIS2 called dBAIS2, which are generated from measurements in the near and middle IR channels of the Sentinel images. The proposed algorithm was tested on Sentinel satellite images acquired from June to September 2021 for the Tizi Ouzou region, Algeria. We found it to have an overall accuracy of 97%, outperforming the results obtained from dNBR and dBAIS2 by large margins.*

## 1 Introduction

Recently, climate change caused by global warming and local factors (human economic activity, land reclamation) have led to an increase in forest fires during the most fire-dangerous spring and summer periods. The large increase in the fire rate has been recorded globally with social impacts (loss of human life), economic effects (damage to houses and infrastructures) and impacts on the climate. Examples are the devastating fires which recently occurred in Europe, Australia (BISHOP 2020; HUGHES et al. 2020; COLLINS et al., 2021), and South America (MARETTI et al. 2014; FRAZER 2019; LIZUNDIA-LOIOLA et al. 2020). Under these conditions, an operational assessment of the consequences of forest fires to plan works to combat fires and restore damaged forest territories acquires significant importance.

The existing satellite monitoring systems for forest fires are based on such thermal sensors as:

1. SEVIRI (Spinning Enhanced Visible and InfraRed Imager) for active fire monitoring.
2. AVHRR (Advanced Very High-Resolution Radiometer) installed on NOAA satellites. This data is used to detect suspected active fires and to generate various cloud masks.
3. MODIS (Medium Resolution Spectroradiometer) installed on the TERRA and AQUA satellites. MODIS images are also used to detect suspected active fires and to build various cloud masks, as well as in the operational assessment of areas covered by fire based on information about active combustion. MODIS data is also used to estimate the areas covered

---

<sup>1</sup> Ukrainian State University of Science and Technologies, Institute of industrial and business technologies, st. Lazaryana, 2, 49010, Dnipro, Ukraine, E-Mail: vvitagat@gmail.com

<sup>2</sup> Dnipro University of Technology, Information Technologies & Computer Engineering Department, av. Dmytra Yavornytskoho, 19, 49005 Dnipro, Ukraine, E-Mail: [Hnatushenko.V.V, Kashtan.v.yu]@nmu.one

<sup>3</sup> Gottfried Wilhelm Leibniz Universität Hannover, Institut fuer Photogrammetrie und GeoInformation, Nienburger Str. 1, D - 30167 Hannover, Germany, E-Mail: heipke@ipi.uni-hannover.de

by fire based on the analysis of the state of vegetation before and after the fires (WOOSTER et al. 2021).

#### 4. VIIRS (visible infrared radiometer kit).

The performance of these systems depends on the accuracy of the algorithms used to detect thermal anomalies. The pros and cons of the algorithms and systems used, including their potential for prompt fire alert, have been analyzed in different studies. For example, AVHRR contains a spectral channel with a wavelength of 2.75  $\mu\text{m}$  (a frequency close to the maximum intensity of infrared radiation from forest fires), as well as channels in the visible and far IR regions of the spectrum. These channels make it possible to distinguish between fires, some atmospheric background, and surface objects that are not distinguishable in the 2.75  $\mu\text{m}$  channel alone. In addition, the AVHRR instrument swath reaches almost 3000 km, albeit due to one of the main shortcomings of this device – the relatively low spatial resolution (at the sub-satellite point, this is about 1  $\text{km}^2$ ).

However, the detection and analysis of forest fires using the above-described sensors without using specially adapted automatic tools is an expensive and difficult process. For this reason, work is currently underway around the world to create automated systems for detecting and evaluating forest fires. At the same time, the main direction in the creation of such systems is the development of methods for the automated processing of satellite data, since this greatly simplifies the process of assessing burnt areas and in some cases (swamps, protected areas) exceeds the accuracy of the assessment on the ground.

## 2 State-of-the-art

Review of literature has shown that a main direction in the development of systems for automated satellite data processing are index-based methods. These are calculated from the spectral channels of images and are sometimes considered to have increased information content about the object of study compared to the original data. An example is the difference normalized burnt ratio (dNBR). For this index, the characteristics of forest vegetation are taken into account: Like other vegetation, it is characterized by high reflectivity in the green part of the spectrum, a minimum of reflection in the blue and red parts, and a sharp increase in a reflection in the near-IR part of the electromagnetic spectrum, while reflectivity in the short wave IR part is again low. This is due to the reflection of green rays and the absorption of blue and red rays by chlorophyll contained in the vegetation, while the high reflectivity in the IR part is due to the transmission of IR rays by chlorophyll and their reflection by the internal tissues of the leaves. Reflection in the IR part is subject to a sharp change due to fire: the IR reflectivity drops and the SWIR value increases. In this regard, in analogy with the well-known Normalized Difference Vegetation Index (NDVI), the NBR is computed as the difference between NIR and SWIR values divided by their sum. A high NBR values indicates unburnt vegetation, while low values indicate burnt areas. For the dNBR determination, the difference between the NBR computed on the date preceding the fire and the date after it, as close as possible to the date of the fire, is determined.

Another popular index is the Burnt Area Index (BAI), which uses the reflectance in the red and the infrared part of the reflectance spectrum to identify areas affected by fire (CHUVIECO et al., 2002). BAIS2, the Burnt Area Index for Sentinel-2, is an extension of BAI (Burned Area Index)



and has been specifically designed to take advantage of the spectral properties of Sentinel-2 imagery (see e.g. KASHTAN & HNATUSHENKO 2023).

Modern approaches to assessing the risk of forest fires, taking into account the influence of natural and anthropogenic environmental factors, differ significantly in different countries. For example, in Australia, the McArthur Forest Fire Danger Index (FFDI) and Forest Fire Behavior Tables (FFBT) are used. Canada, several US states, Europe, Mexico, New Zealand, and the countries of Southeast Asia use the Canadian Forest Fire Weather Index (FWI). The United States has developed and operates the National Fire Danger Rating System (NFDRS). In addition, as indicated in (MAHMOUD et al. 2018), Wildland – Urban Interface (WUI) maps in the United States cover more than two decades of data. However, they do not investigate the influence of this factor on the effectiveness of early detection of fires from space. However, given the variety of factors that affect the possibility of a forest fire, an integrated approach including both, natural and anthropogenic factors, is urgently needed.

Along another line of research, GIGLIO et al. (2016) have proposed an improved lane-level fire detection algorithm that has been implemented as part of the processing of Collection 6 ground products, updating the algorithm to eliminate limitations of the Collection 5 product, namely the occurrence of false alarms caused by small forest areas and the decrease of large fires obscured by thick smoke.

Recently, convolutional neural networks have been successfully used in image recognition and have achieved higher accuracy than traditional methods for object recognition and in semantic segmentation (for work of the authors, see, e.g., HNATUSHENKO & ZHERNOVYI 2019, 2020; HEIPKE & ROTTENSTEINER 2020). Due to the layered structure, the neural network can implement an approximation of complex functions. A convolutional neural network reduces the number of parameters by locally sharing the weights to improve learning efficiency.

Deep learning methods are also increasingly being used for forest fire detection from images. ZHAO et al. (2022) proposed an advanced Fire-YOLO deep learning algorithm for detecting fire and smoke details of forest fires in images. The Fire-YOLO model effectively copes with the search for small fire sites. The methods proposed in (MUHAMMAD et al. 2018; HE et al. 2021) are based on convolutional neural networks for fire detection (TRAN et al. 2018) and for smoke detection based on feature-level and decision-level fusion (KEY et al. 2002). These methods require relatively high computational power.

### 3 Methodology

This work aims to improve the efficiency of monitoring a forest fire based on Sentinel-2 images and deep learning. The Sentinel-2A and Sentinel-2B satellite data yield the best compromise of spatial, spectral, and temporal indicators among publicly available satellite data today, which is especially important given the dynamics of natural processes. In terms of revisit time, the Sentinel-2A and Sentinel-2B satellites are more than three times superior to the Landsat 7, and Landsat 8 satellites, making repeated surveys of the same area of the earth's surface possible after 5 days instead of 16 and 18 days, respectively.

Figure 1 presents a block diagram of the proposed method. It consists of the following steps:

Step 1. Selection of data in the form of satellite image bands. A satellite image of any size, including data from IR channels, serves as the input of our system.

Step 2. Preprocessing. It is necessary to eliminate various distortions which occurred during data acquisition. Preprocessing consists of interference filtering, radiometric and geometric data correction incl. spectral correction (HNATUSHENKO & KASHTAN 2021).

Step 3. CNN training and classification. Three classes were defined form the classification: fire, combined with burned areas, called *fire-burned* here (i.e. current fires and areas which were recently burned), *smoke* and *background*. As a result, we have the probabilities of each pixel to belong to one of those classes. The final label is taken to be the class with the maximum probability estimate among the two classes and background.

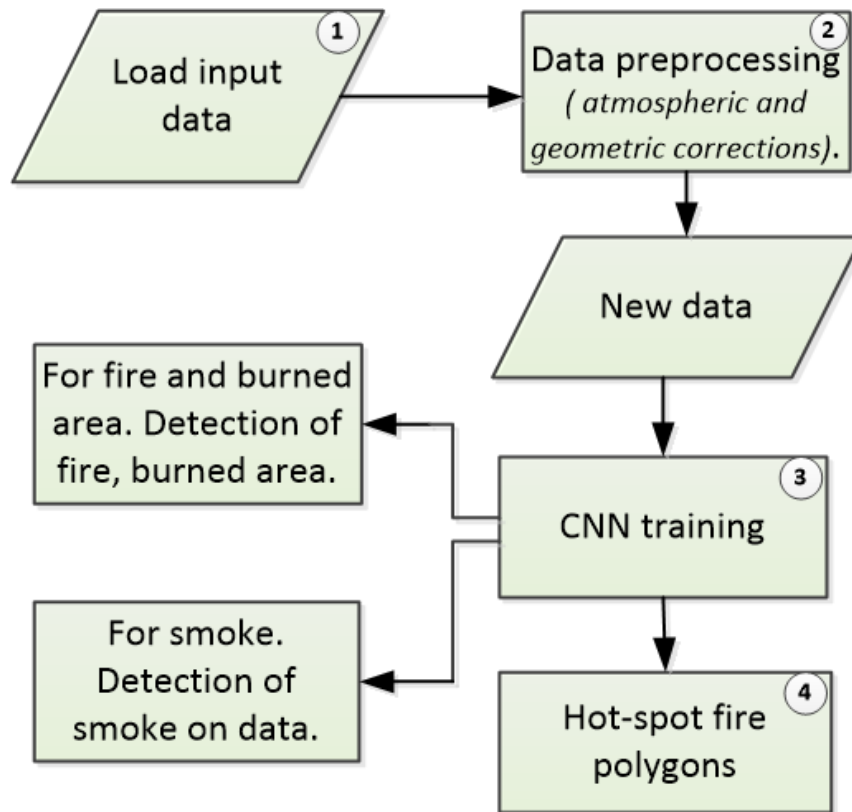


Fig. 1: Algorithm for detecting fires and forest burnt areas

This paper proposes to use a U-Net structure of a convolutional neural network (RONNEBERGER et al. 2015) that combines convolution and subsampling layers, as shown in Figure 2. A satellite image goes through sequential convolution operations with a kernel size of  $5 \times 5$ , and  $3 \times 3$ , respectively. A  $3 \times 3$  subsampling layer with a stride of 3 follows the second and fourth convolutional layers and then, a final subsampling layer with a  $1 \times 1$  kernel size with a stride of 1 is applied. One of the important stages of a neural network is the choice of the activation function. The classic backpropagation algorithm works fine on neural networks with two and three layers, but as the depth of the map increases, difficulties appear (for example, attenuation of gradients). This work proposes to use a ReLU type function. One of the advantages of ReLU over other

activation functions is that it does not fire all neurons together. Some disadvantage of ReLU is, however, that the function value in the negative region is zero. For example, a large gradient passing through a ReLU may cause the weights to be updated such that the given neuron is not activated, and thus, the gradient passing through this neuron will be zero. If the gradient is zero, none of the weights will be update during back propagation (KASHTAN & HNATUSHENKO 2023). To offset this shortcoming, we use Leaky ReLU (LReLU) (XU et al., 2015; GOTTHANS et al. 2022) in our work. While the usual ReLU gives zero output in the interval  $x < 0$ , LReLU returns a small negative value on this interval. That is, the function for LReLU has the form:

$$f(x) = \alpha x \text{ for } x < 0 \text{ and } f(x) = x \text{ for } x \geq 0, \quad (1)$$

where  $\alpha$  is a small positive constant (about 0.01). We use the well-known SoftMax function as loss function.

Step 4. Building of hotspot fire polygons. To identify hot spots, we generate connected components for the class *fire-burned*. We can thus retrieve polygons of the respective areas with various attributes: position, size, etc.

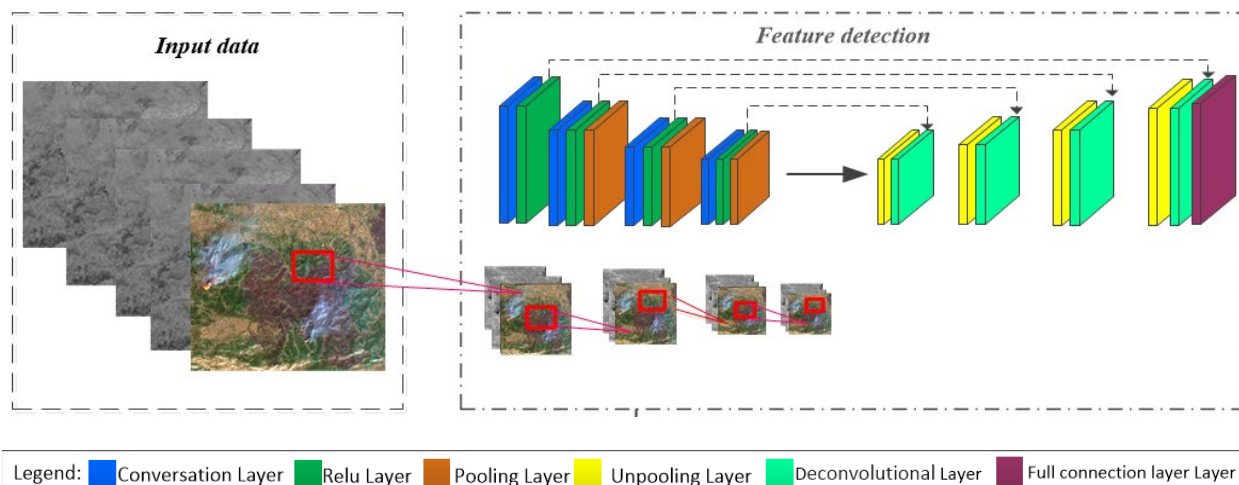


Fig. 2: Structure of the proposed CNN for fire detection

## 4 Experimental results for index based methods

In order to be able to put the results of the proposed CNN method into context we first report what was obtained with more traditional index-based methods; we have chosen the two indices dNBR and dBAIS2. The test area for carrying out the experiments is the Tizi Ouzou region in Algeria (Figure 3). A fire was identified from Sentinel-2 satellite imagery acquired in 2021, taken between July and the end of September. The images were taken during a burning fire with a cloud of smoke, which complicates data processing.

The results using the indices dNBR and dBAIS2 are shown in Figure 4. The spectral index of the burnt area for Sentinel-2 was calculated from the difference between the BAIS2 values before and after the fire, which we call dBAIS2. As can be seen from Figure 4, the smallest values of dNBR and dBAIS2 in the range of 0–50 are typical for areas that have undergone the most severe burning,

or areas with open fire. The area with smoke from the fire is characterized by values close to the maximum (about 250). The unburnt area is characterized by the homogeneous values and their consolidation near the 200 mark. It can be visually seen that the spectral indices are not ideal, since there are pixels in the range of 0–50 (red color) on satellite images where there was no fire. Moreover, to obtain correct values, it is necessary to apply a water mask. This excludes water pixels and increases the accuracy of mapping the fire site.



Fig. 3: Satellite image after fire, Sentinel 2 Band4-Band3-Band2, from Tizi Ouzou

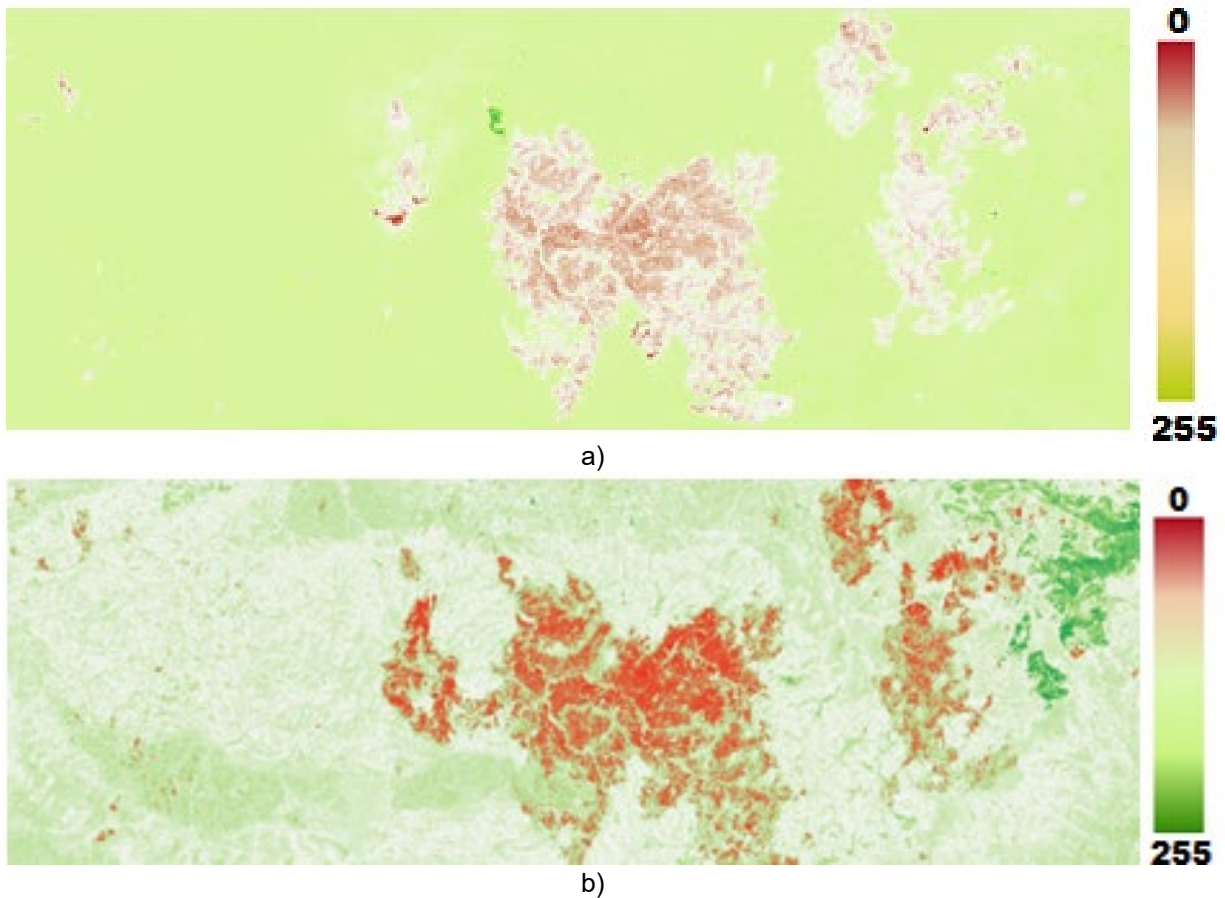


Fig. 4: Calculation index indicator results from Tizi Ouzou region (12.08.2022): a) dNBR; b) dBAIS2

## 5 Results with the CNN method and accuracy assessment

Given enough experience with the problem at hand and the availability of standards of burnt areas, visual methods for identifying burnt areas are often more accurate than automated algorithms. Based on this expert information, this ground truth was collected and subsequently compared to the automatically derived results in order to assess the quality of the proposed method for automated detection of forest fire consequences based on CNNs, and to compare the results to those obtained with the index-based methods. In the course of the work, a number of satellite images was selected, in which it was possible to visually distinguish and delineate the different relevant classes, namely *fire-burned*, *smoke* and *background*. The total number of training data was 180 images. The total number of tested images in the set was 100 images with a size of 1266×484 pixels each. We used 4, 3, 2, 12, 8A (Red, Green, Blue, SWIR, Narrow NIR) Sentinel-2 image channels for the classification, only post-fire images were used.

To evaluate the accuracy of the classification a confusion matrix of the true positive (TP, the classifier and the expert both detect fire in the image), true negative (TN, the classifier and the expert both detect no-fire areas), false positive (FP, the classifier detects fire, while the expert does not), and false negative (FN, the classifier does not detect fire, while the expert does) has been calculated by using the Tizi Ouzou region dataset (KAPLAN & AVDAN 2018). Not that we combined the classes *smoke* and *background* into *non-burned* as we are mainly interested which areas were effected by fire.

More details can be derived from the confusion matrix (see Table 1, shown for the CNN method only).

Tab.1: Confusion matrix for CNN method, Tizi Ouzou region

Fire status		Fire-burned	Non-burned
		Expert result	
Algorithmic results	Fire-burned	TP = 96.8 %	FP = 3.2 %
	Non-burned	FN = 2.5 %	TN = 97.5 %

From the confusion matrix the overall accuracy can be computed according to

$$\text{Overall Accuracy} = (TP + TN) / (TP + TN + FP + FN). \quad (2)$$

For the CNN method the overall accuracy according to eq. (2) for the Tizi Ouzou region is 97%. The overall accuracy for the index-based methods were found to be significantly less, i.e. 82.1% for dNBR and 87.3% for dBAIS2 (Table 2).

Tab. 2: Accuracy assessment of the spectral indicators and the proposed method

Territories	Overall accuracy, %		
	dNBR	dBAIS2	<b>CNN method</b>
Tizi Ouzou region	82.1	87.3	<b>97</b>

Figure 5 shows a map of fire polygons for Tizi Ouzou. The automatically derived data are shown in yellow, the ground truth data in green.

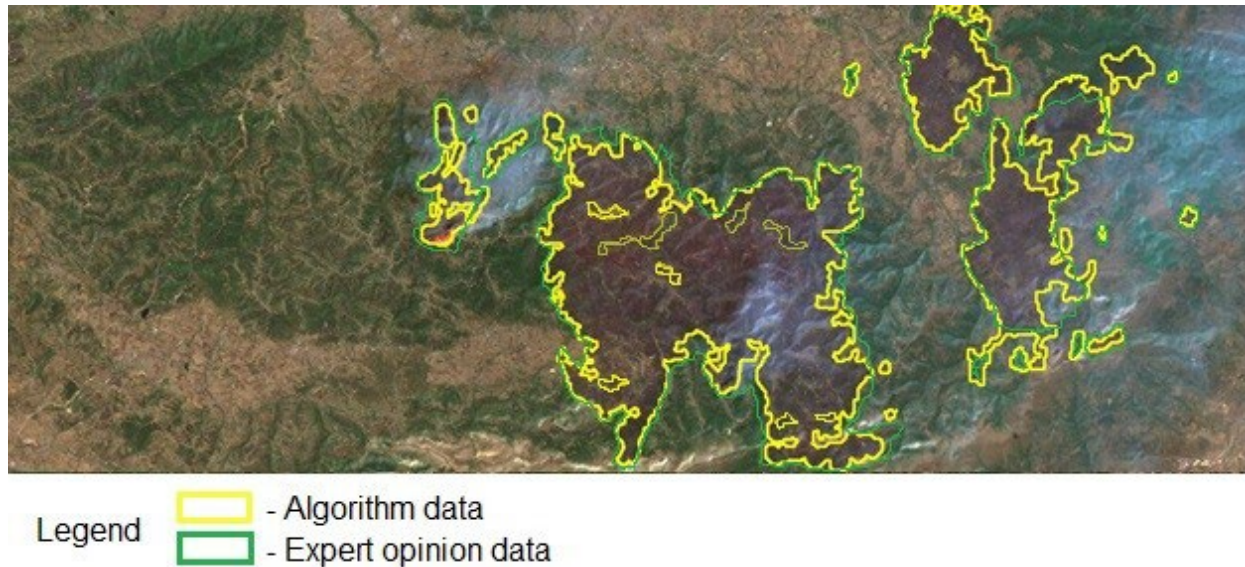


Fig. 5: Map of the territories covered by the fire from Tizi Ouzou

## 6 Conclusions & Outlook

Nowadays, the difference normalized burnt ration (dNBR) one of the most popular tools for analyzing burnt areas and fires in existing systems for detecting and evaluating fires, however the method is based on the spectral values of individual pixels without considering the neighborhood, and is thus rather error-prone.

This paper presented a method for automated fire detection from Sentinel-2A and Sentinel-2B satellite data, yielding the currently best compromise between spatial, spectral, and temporal resolution among publicly available satellite data, with convolutional neural networks. The functionality of the created system allows solving the task, starting from the moment of receiving the input data and ending with the export of a hot-spot fire polygonal file describing the area that was burnt. For the Tizi Ouzou, Algeria region, an overall accuracy of 97% was achieved.

Our future research will be devoted to generalizing these results to more and larger test sites and to training a neural network to classify the degree of fire impact (detection of the degree of damage to forest stands), which will require field validation. Another direction of research will be the consideration of the often-unbalanced dataset, e.g. by altering the loss function to give more emphasize to classes with fewer pixels.

## 7 References

- BISHOP, J., 2020: Burnt Assets: The 2019–2020 Australian Bushfires; WWF: Sydney, Australia, 34, <https://doi.org/10.13140/RG.2.2.23413.50404>.
- CHUVIECO, E., MARTIN, M. P. & PALACIOS, A., 2002: Assessment of Different Spectral Indices in the Red-Near-Infrared Spectral Domain for Burnt Land Discrimination. *International Journal of Remote Sensing*, **23**(23), 5103-5110, <https://doi.org/10.1080/01431160210153129>.
- COLLINS, L., BRADSTOCK, R. A., CLARKE, H., CLARKE, M. F., Nolan, R. H. & PENMAN, T. D., 2021: The 2019/2020 mega-fires exposed Australian ecosystems to an unprecedented extent of high-severity fire. *Environ. Res. Lett.* 16, <https://doi.org/10.1088/1748-9326/abeb9e>.
- FRAZER, J., 2019: The Origin of Fire in South America. In: *Myths of the Origin of Fire*, 123-135, <https://doi.org/10.4324/9781315028798-11>.
- GIGLIO, L., SCHROEDER, W. & JUSTICE, C., 2016: The Collection 6 MODIS active fire detection algorithm and fire products. // *Remote Sensing of Environment*, **178**, 31-41, <https://doi.org/10.1016/j.rse.2016.02.054>.
- GOTTHANS, J., GOTTHANS, T. & MARSALEK, R., 2020: Deep convolutional neural network for fire detection, 2020 30th International Conference Radioelektronika, 5-16 April 2020, Bratislava, Slovakia, 1-6.
- HE, L., GONG, X., ZHANG, S., WANG, L. & LI, F., 2021: Efficient attention based deep fusion CNN for smoke detection in fog environment, *Neurocomputing*, **434**, 224-238.
- HEIPKE C. & ROTTENSTEINER F., 2020: Deep learning for geometric and semantic tasks in photogrammetry and remote sensing. *Geo-spatial Information Science*, **23**(1), 10-19, <https://doi.org/10.1080/10095020.2020.1718003>.
- HNATUSHENKO, V. & KASHTAN V., 2021: Automated pansharpening information technology of satellite images. *Radio Electronics, Computer Science, Control.*, **2**, P.123-132. <https://doi.org/10.15588/1607-3274-2021-2-13>.
- HNATUSHENKO, V. & ZHERNOVYI, V., 2019: Complex Approach of High-Resolution Multispectral Data Engineering for Deep Neural Network Processing. *Lecture Notes in Computational Intelligence and Decision Making*, 659-672, [https://doi.org/10.1007/978-3-030-26474-1\\_46](https://doi.org/10.1007/978-3-030-26474-1_46).
- HNATUSHENKO, V. & ZHERNOVYI, V., 2020: Method of Improving Instance Segmentation for Very High Resolution Remote Sensing Imagery Using Deep Learning. In: Babichev S., Peleshko D., Vynokurova O. (eds) *Data Stream Mining & Processing. DSMP 2020. Communications in Computer and Information Science*, **1158**, Springer, Cham. [https://doi.org/10.1007/978-3-030-61656-4\\_21](https://doi.org/10.1007/978-3-030-61656-4_21).
- HUGHES, L., STEFFEN, W., MULLINS, G., DEAN, A., WEISBROT, E. & Rice, M., 2020: *Summer of Crisis*; Climate Council of Australia Limited: Potts Point, Australia, 34.
- KAPLAN, G. & AVDAN, U., 2018: Sentinel-1 and Sentinel-2 Data fusion for wetlands mapping: Balıkdami, Turkey. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, **42**(3), 729-734, <https://doi.org/10.5194/isprs-archives-XLII-3-729-2018>.

- KASHTAN, V. & HNATUSHENKO, V., 2023: Deep Learning Technology for Automatic Burnt Area Extraction Using Satellite High Spatial Resolution Images. In: Babichev, S., Lytvynenko, V. (eds) *Lecture Notes in Data Engineering, Computational Intelligence, and Decision Making. ISDMCI 2022. Lecture Notes on Data Engineering and Communications Technologies*, **149**. Springer, Cham. [https://doi.org/10.1007/978-3-031-16203-9\\_37](https://doi.org/10.1007/978-3-031-16203-9_37).
- KEY, C., BENSON, N., OHLEN, D., HOWARD, S., MCKINLEY, R. & ZHU, Z., 2002: The normalized burn ratio and relationships to burn severity: ecology, remote sensing and implementation. In: Greer JD, editor. *Rapid delivery of remote sensing products. Proceedings of the Ninth Forest Service Remote Sensing Applications Conference*, 8-12.
- LIZUNDIA-LOIOLA, J., PETTINARI, M. & CHUVIECO, E., 2020: Temporal Anomalies in Burnt Area Trends: Satellite Estimations of the Amazonian 2019 Fire Crisis. *Remote Sens.*, **12**, 151.
- MAHMOUD, K., MRIGANK, R., NOMAN, M., NEIL D., YANG W. & FARKHUND I., 2018: Malware Classification with Deep Convolutional Neural Networks. 2018 9th IFIP International Conference on New Technologies, Mobility and Security (NTMS). <https://doi.org/10.1109/NTMS.2018.8328749>.
- MARETTI, C.C., HOFSTED, R. & GRANIZO, T., 2014: State of the Amazon: Ecological Representation in Protected Areas and Indigenous Territories. *WWF Living Amazon Initiative - State of the Amazon*.
- MUHAMMAD, K., AHMAD, J. & BAIK, S., 2018: Early fire detection using convolutional neural networks during surveillance for effective disaster management. *Neurocomputing*, **288**, 30-42.
- RONNEBERGER, O., FISCHER, P. & BROX T., 2015: U-Net: Convolutional networks for biomedical image segmentation. 18<sup>th</sup> Int. Conference on medical image computing and computer assisted intervention. LNCS, **9351**, Springer, 234-241.
- TRAN, B., TANASE, M., BENNETT, L. & APONTE, C., 2018: Evaluation of spectral indices for assessing fire severity in Australian temperate forests. *Remote Sensing*, **10**(11):1680. <https://doi.org/10.3390/rs10111680>.
- WOOSTER, M., ROBERTS, G., GIGLIO, L., ROY, D., FREEBORN, P., BOSCHETTI, L., JUSTICE, C., ICHOKU, C., SCHROEDER, W., W., BOSCHETTI, L., DAVIES, D.K., SMITH, A.M.S., CSISZAR I., FROST, P., SETZER, A., ZHANG, T., de JONG, M.C., JOHNSTON, J. M., Xu, W., SCHMIDT, C., STRYDOM, T., ELLISION, L., VADREUVU, K.P., TANPIPAT, V., SAN-MIGUEL-AYANZ, J., SPARKS, A.M., MCCARTY, J. L. & NGUYEN, H., 2021: Satellite remote sensing of active fires: History and current status, applications and future requirements. *Remote Sens. Environ*, **267**, 112694, <https://doi.org/10.1016/j.rse.2021.112694>.
- XU, B., WANG, N., CHEN, T. & LI, M., 2015: Empirical Evaluation of Rectified Activations in Convolutional Network. <https://doi.org/10.48550/arXiv.1505.00853>.
- ZHAO, L., ZHI, L., ZHAO, C. & ZHENG, W., 2022: Fire-YOLO: A Small Target Object Detection Method for Fire Inspection. *Sustainability*, **14**, 4930, <https://doi.org/10.3390/su14094930>.



# Qualitätsuntersuchungen UAV-getragener Kamerasysteme am Testfeld Bochum

**ANSGAR GREIWE<sup>1</sup>, VOLKER SPRECKELS<sup>2</sup>, ANDREAS SCHLIENKAMP<sup>2</sup>,  
RAINER BRECHTKEN<sup>1</sup> & SIMON DAMBERG<sup>1</sup>**

*Zusammenfassung: Oberflächenmodelle können zum Beispiel aus diskreten stereoskopischen Messungen, Bildkorrelation oder Structure from Motion (SfM) erzeugt werden. Im photogrammetrischen Normalfall werden per stereo-Betrachtung und der Bildkorrelation die Genauigkeiten durch Kenngrößen der Bündelblockausgleichung nachgewiesen. Für einen Genauigkeitsnachweis der automatisierten SfM-Verfahren reicht diese klassische Betrachtungsweise allerdings nicht aus. Hier hängt das Dense Matching zusätzlich von der Bildqualität der eingesetzten Sensorik ab. Somit ist neben den Ergebnissen der Bündelblockausgleichung ein Vergleich mit Referenzflächen oder Referenzgeometrien erforderlich.*

*Dieser Beitrag ist zweigeteilt, denn über eine Untersuchung der optischen Eigenschaften eines eingesetzten Kamerasystems können bereits vorab im Labor Vorhersagen zur Qualität einer resultierenden Punktwolke getroffen werden. Diese werden dann im Praxiseinsatz betrachtet, wobei drei UAV-getragene Kamerasysteme am neu eingerichteten UAV-Testfeld der Hochschule Bochum untersucht und deren Einfluss auf die erzeugte Punktwolke bzw. das Oberflächenmodell dargestellt wird.*

## 1 Hintergrund, Motivation, Problemstellung

Werden digitale Oberflächenmodelle zur Deformationsüberwachung einer Geländeoberfläche eingesetzt, so sind klare Genauigkeitsvorgaben gegeben, insbesondere im Hinblick auf die Höhengenaugigkeit der erzeugten Punkte (SPRECKELS et al. 2016).

Oberflächenmodelle können zum einen aus diskreten Messungen, wie zum Beispiel aus einer stereoskopischen Punktmessung erzeugt werden. Eine weitere Möglichkeit besteht in der automatisierten Generierung des Oberflächenmodells durch Structure from Motion (SfM), hier insbesondere per Dense Image Matching. Im Falle der zuerst genannten manuellen Bestimmung von Oberflächenpunkten wird der eingangs geforderte Genauigkeitsnachweis durch Kenngrößen der Bündelblockausgleichung nachgewiesen. Hierzu gehören die Angaben zu den erreichten Standardabweichungen an Pass- und Kontrollpunkten, bei vorliegender oder im Felde durchgeführter Kamerakalibrierung zudem die Standardabweichungen der inneren Orientierung. Die Standardabweichung eines (manuell gemessenen) Kontrollpunktes wird dann ebenso für die zu messenden Neupunkte verwendet.

Im zweiten Fall, der Ableitung von Oberflächenpunkten durch Dense Image Matching, repräsentiert dieser Genauigkeitsnachweis jedoch nicht Qualität (Rauschen) und Auflösung

---

<sup>1</sup> Hochschule Bochum, Labor für Optische 3D-Messtechnik,  
Am Hochschulcampus 1, D-44801 Bochum, E-Mail: Ansgar.Greiwe@hs-bochum.de

<sup>2</sup> RAG Aktiengesellschaft, Unternehmensbereich Standort-/Geodienste, K-SG-NG Geodaten-Fernerkundung, Im Welterbe 10, D-45141 Essen

(Detailreichtum) der erzeugten Punktwolke. Diese Parameter sind in erster Linie von der Zuverlässigkeit des eingesetzten Interest Operators abhängig - und dieser wiederum von der Bildqualität. So zeigen GREIWE et al. (2017), dass die Ableitung einer Geländeoberfläche aus Bilddaten durch SfM im hohen Maße von der Bildqualität des verwendeten Kamerasystems abhängt.

## 2 Kontrastauflösung als Maß für die Bildqualität

Die Leistungsfähigkeit – und damit die Bildqualität - eines optischen Systems ist für den Betrachter immer dann offensichtlich, wenn scharfe Kanten oder feine Strukturen in einer Aufnahme unscharf erscheinen. Werden äußere Einflüsse wie die Bewegungsunschärfe bei Verwendung des Sensors auf einer bewegten Plattform ausgeschlossen, so sind entweder Abbildungsfehler der Optik (Aberration, Astigmatismus) oder das Auflösungsvermögen des eingesetzten Sensors die limitierenden Faktoren für eine scharfe Abbildung (NASSE 2008).

Das Auflösungsvermögen – und damit die optische Leistungsfähigkeit eines digitalen Kamerasystems – lässt sich durch die Bestimmung der Modulationstransferfunktion (MTF) bestimmen (CUNNINGHAM & FENSTER 1987; REICHENBACH et al. 1991). Hierbei wird die Modulation als Quotient des Kontrastes eines Signals zum Kontrast im Bild ermittelt. Eine Modulation von 10% entspricht ungefähr dem Auflösungsvermögen des menschlichen Auges, einen Hell-Dunkel-Wechsel von „strukturlosem Grau“ zu unterscheiden (NASSE 2008).

Die MTF gibt den Verlauf der Modulation über die Ortsfrequenz eines Signals an. Die Messung der MTF kann durch die Verwendung von Siemens-Stern-Charts erfolgen (MEIßNER et al. 2018). Die sich ändernde Ortsfrequenz wird hierbei durch die stetig breiter werdenden Strahlen des Siemenssterns realisiert. Die Ermittlung der MTF kann jedoch nur an den Stellen im Bild erfolgen, in denen auch ein Siemensstern platziert ist. MEIßNER et al. (2018) ermitteln so für unterschiedliche Kameras die MTF im Zentrum und über eine Hauptdiagonale im aufgenommenen Bild (Abb. 1).

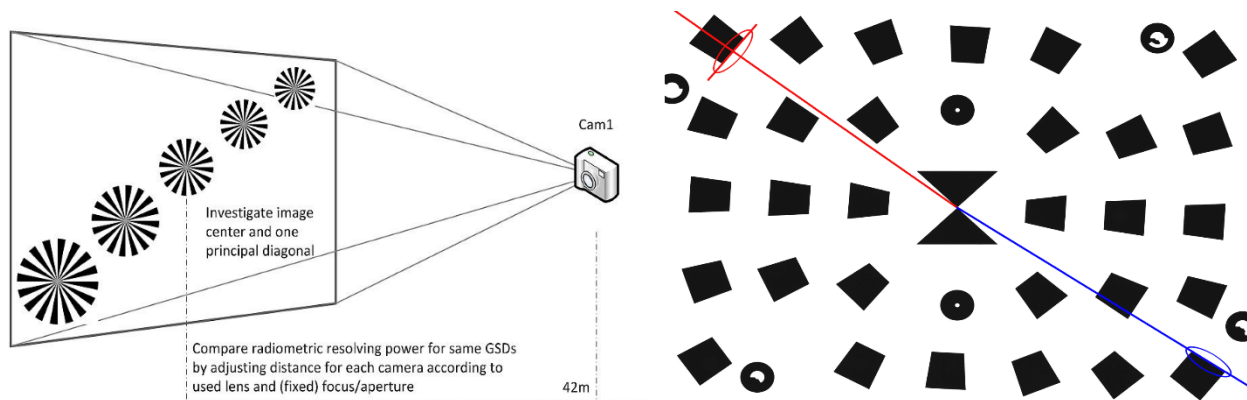


Abb. 1: Testcharts zur Bestimmung der MTF-Kurve: Siemensstern (links aus MEIßNER et al. 2018) und Slant-Edge Testchart (rechts, eigene Aufnahme)

Eine alternative Methode ist die Verwendung eines "Slant-Edge"-Charts nach ISO 12233. Hierbei wird der Kontrast an schräg verlaufenden Linien auf einem homogenen Hintergrund ermittelt. Die Änderung der Ortsfrequenz des Signals wird durch die schrägen Linien realisiert, welche die Pixel

des Sensors in stetig steigendem Maße schneiden. Der Vorteil dieser Methode liegt in der Möglichkeit, die MTF flächenhaft über die gesamte Sensorfläche mit einer einzigen Aufnahme zu ermitteln, da als Testmuster deutlich mehr Kanten als Siemenssterne auf einem Testchart platzierbar sind.

Die MTF-Messung in Richtung des Bildradius (sagittal) erfolgt an den in Abbildung 1 blau dargestellten Kanten, die tangentiale (meridionale) MTF an den rot hervorgehobenen Kanten. Die Ortsfrequenz kann für eine MTF durch den Quotienten Linienpaare pro Millimeter (lp/mm) angegeben werden. Je größer der Quotient, desto höher die Ortsfrequenz resp. desto feiner die Strukturen im Bild. So kann für eine gegebene Ortsfrequenz der Verlauf des Auflösungsvermögens über die Sensorhöhe (Bildradius) dargestellt werden (Abb. 2).

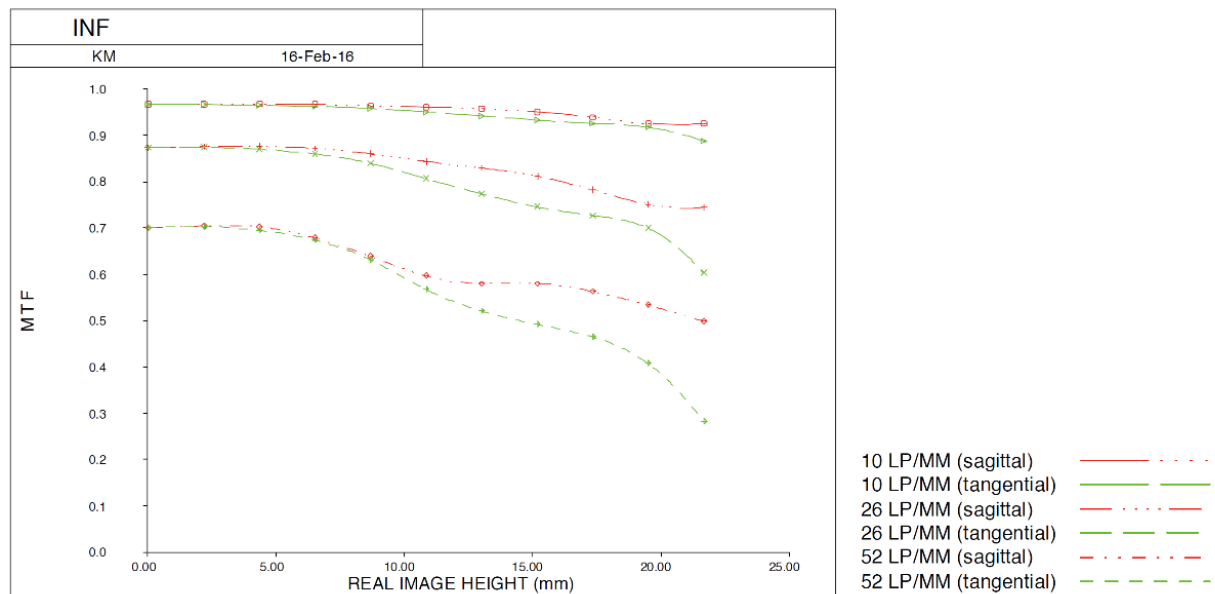


Abb. 2: MTF-Kurve des Sensors P1 (DJI) mit Objektiv DL 50 mm F2.8 LS (aus: DJI 2021)

Bei einer geringen Ortsfrequenz von 10 lp/mm ist über das gesamte Bild eine hohe Kontrastauflösung zu beobachten. Wird eine hohe Ortsfrequenz mit 52 lp/mm vorgegeben, so lassen sich zwei Effekte beobachten: Zum einen nimmt die Kontrastauflösung wie erwartet zum Bildrand hin ab, zum anderen ist zwischen einer sagittalen und tangentialen (meridionalen) Bestimmung der MTF ein Unterschied erkennbar.

Mit dieser Betrachtung lassen sich Rückschlüsse auf die Leistungsfähigkeit von Objektiven zu, was im Folgenden vorgestellt wird.

## 2.1 Fokus

Zunächst beeinflusst die Fokuseinstellung eines Objektivs die MTF eines Kamerasystems (siehe BOREMANN 2001). In einem Laborversuch konnte dieser Effekt dadurch nachgewiesen werden, dass das Testchart außerhalb des Schärfentiefebereiches der eingesetzten Vollformatkamera positioniert wurde. Bei kleinformatischen Sensoren wird die Verwendung eines Autofokus zur photogrammetrischen Aufnahme als „zuverlässig und gut“ bezeichnet (ZUHORST & HUDASCH 2020). Bei der Verwendung größerer Voll- oder Mittelformatsensoren sollte durch Verwendung

eines manuellen Fokus diese Fehlerquelle aber ausgeschlossen und zudem der Schärfentiefebereich beachtet werden. So besitzt die Ausrüstung von GHASSOUN et al. (2021) bestehend aus einer Mittelformatkamera und einem Objektiv mit 150 mm Brennweite eine Schärfentiefe von 14 m bei einer fokussierten Dingweite von 35 m.

## 2.2 Datenformat

UAV-getragene Kamerasysteme bieten dem Nutzer unter Umständen mehrere Möglichkeiten zur Speicherung der Bilddaten. Das RAW-Format zeichnet die Daten verlustfrei auf, es ist jedoch proprietär und bedarf einer digitalen Entwicklung in ein anderes, verlustfreies Datenformat (z.B. TIFF), bevor die Daten für die Auswertesoftware lesbar sind. Andere Systeme führen eine kamerainterne Prozessierung der Bilddaten und die Speicherung im JPEG-Format durch. Die Prozessierung in das verlustbehaftete JPEG-Format führt jedoch zu Artefakten und einer messbaren Verschlechterung der Bildqualität. Es sollte also für die Ableitung von Oberflächenmodellen vermieden werden (VERHOEVEN 2010): „*For strictly scientific photography, such as airborne remote sensing, shooting RAW is mandatory practice because it is the only way to assess the originally captured DN's, [...]*“. Einige Kamerasystemhersteller lassen aus diesem Grund keine Speicherung im JPEG zu, sollte sie dem Nutzer angeboten werden, ist sie zu deaktivieren.

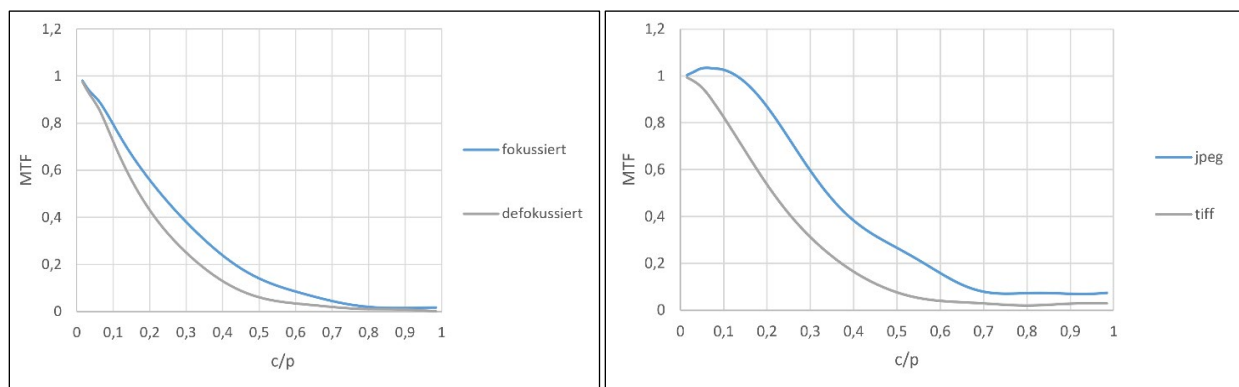


Abb. 3: Einfluss von Fokus (links) und Datenformat (rechts) auf die MTF. Im Gegensatz zur Angabe lp/mm wird hier die MTF in Cycle/Pixel (c/p) angegeben. Eine Ortsfrequenz in Pixelgröße (Nyquist) entspricht 0,5 c/p.

Oft wird bei einer kamerainternen JPEG-Konvertierung auch eine Nachschärfung der Bilddaten vorgenommen, die neben den Kompressionsartefakten über diesen Bildverarbeitungsschritt zu weiteren unerwünschten Bildinhalten führt. Besonders deutlich zu erkennen ist dieses Phänomen an signalisierten Passpunkten, denn das Schachbrett erscheint nun an den Kanten besonders dunkel oder hell. Diese Artefakte erzeugen MTF-Werte über 100% (die Aufnahme ist schärfer als das Signal, Abb. 3 rechts).

## 2.3 Blende

Je nach Güte des Objektivs treten unterschiedliche Abbildungsfehler auf, die das Kontrastaufklärungsvermögen beeinflussen. Hierbei kann herstellerseitig die chromatische Aberration durch spezielle Linsenkonstruktionen unterschiedlicher Dispersion (achromatischer

oder dichromatischer Aufbau) begrenzt werden. Durch Abblenden verringern sich die Einflüsse der sphärischen Aberration, bei der achsferne Strahlen keinen gemeinsamen Brennpunkt besitzen, sowie des Astigmatismus, bei dem horizontale und vertikale Strahlen zwei Brennlinien anstelle eines Brennpunktes besitzen.

Im Rahmen dieser Untersuchung sind diese Effekte in einer Messreihe mit einer Sony Alpha 7 R IV für drei Objektive nachgewiesen worden.

Hierbei wurde jeweils eine Blendenreihe unter identischen Laborbedingungen durchgeführt, für zwei Festbrennweitenobjektive (Zeiss Loxia 50 mm und Sony FE 50 mm F1.8) sowie für ein Zoom-Objektiv (Sony FE 24-70 mm F2,8), eingestellt auf 50 mm Brennweite. Für diese drei Objektive wurde jeweils eine Blendenreihe aufgenommen und die MTF flächig über den gesamten Bildbereich sowohl sagittal als auch meridional bestimmt (Abb. 4). Die dargestellten MTF-Kurven zeigen deutlich, dass die Objektive für die Blenden  $f/5.6$  (Zeiss) bzw.  $f/8$  (Sony) optimiert sind.

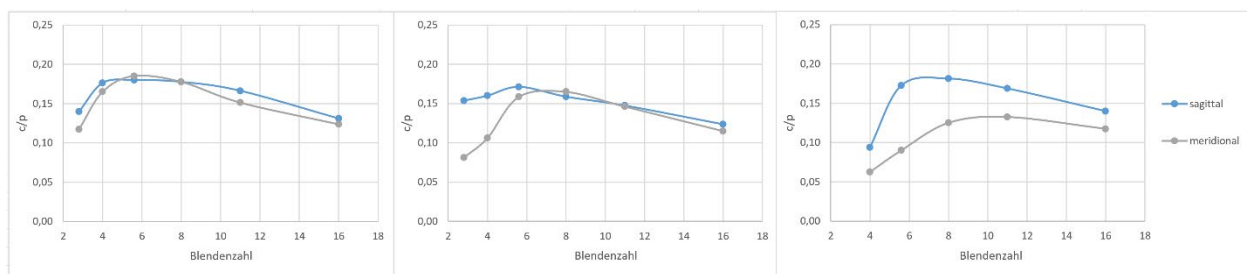


Abb. 4: MTF-Kurven Zeiss Loxia (links), Sony FE-50 mm (Mitte) und Sony FE 24-70 mm (rechts).

Zudem lässt sich aus den unterschiedlichen Messergebnissen in sagittaler und meridionaler Richtung der Einfluss des Astigmatismus ableiten. Das Zoom-Objektiv in Abbildung 4 rechts weist die geringste Bildqualität auf, die als Folge des Optikdesigns über die unterschiedlichen Brennweiten gewertet werden kann. Die Aussage, dass „[...] moderne Optiken auch bei nahezu voller Blendenöffnung gute Leistungsdaten bieten.“, (PRZYBILLA 2022), ist somit nicht allgemeingültig. Je nach Güte der Objektive sind Aufnahmen mit Offenblende zu vermeiden, um die Abbildungsfehler insbesondere am Bildrand zu verringern. Einige Hersteller erlauben in der Kameraansteuerung aus diesem Grund lediglich  $f/5.6$  als größte Blendenöffnung.

## 2.4 ISO

Bewegungsunschärfen während der Aufnahme wirken sich negativ auf die Modulationsübertragung aus. Die tolerierbare Bildwanderung  $\Delta s'_{max}$  kann als Funktion des Auflösungsvermögens  $AV$  (in lp/mm) eines Sensors ausgedrückt werden und wird für Anwendungen in der Aerophotogrammetrie nach LUHMANN (2018) vorgegeben durch:

$$\Delta s'_{max} = 1.5 AV^{(-1)} \quad (1)$$

Die maximal zulässige Bildwanderung der DJI P1 mit einem Auflösungsvermögen von 52 lp/mm (MTF 30% am Bildrand, Abb. 2) beträgt demnach 2,88  $\mu\text{m}$ , was ungefähr 0,66 Pixel entspricht. Bei einem GSD von 9 mm und einer Fluggeschwindigkeit von 2 m/s ist die Belichtungszeit also kürzer 1/333 s zu wählen.

Die Phase One iXU erreicht ein Auflösungsvermögen von 104 lp/mm bei vergleichbarer MTF (Kap. 2.5) und verlangt somit 1/640 s um Bewegungsunschärfen während des Fluges bei 2 m/s zu vermeiden.

Wird die Blendeneinstellung fixiert, kann bei wechselnden Beleuchtungsverhältnissen durch die variable Wahl der Lichtempfindlichkeit des Sensors (ISO) eine korrekte Belichtung sichergestellt werden. In der Praxis wird dieses Vorgehen insbesondere bei Kamerasystemen mit ausschließlicher JPEG-Kompression (s. Kap. 2.2) durchgeführt.

Oft reicht die visuelle Beurteilung einer Aufnahme mit hoher ISO nicht aus, um Einstellungen bis ISO 800 zuzulassen (PRZYBILLA 2022). Um diesen Effekt zu untersuchen, wurden MTF-Testcharts aufgenommen, wobei eine aktuelle Kamera mit geringem Rauschverhalten Sony Alpha 7 R IV mit dem Zeiss Loxia Objektiv eingesetzt wurde. Variiert wurden ISO und Belichtungszeit. Insbesondere in hochfrequenten Bilddetails, also feinen Strukturen mit  $< 0.5$  c/p, wird bereits ab ISO 200 ein messbarer Einfluss auf die Bildqualität festgestellt (Abb. 5).

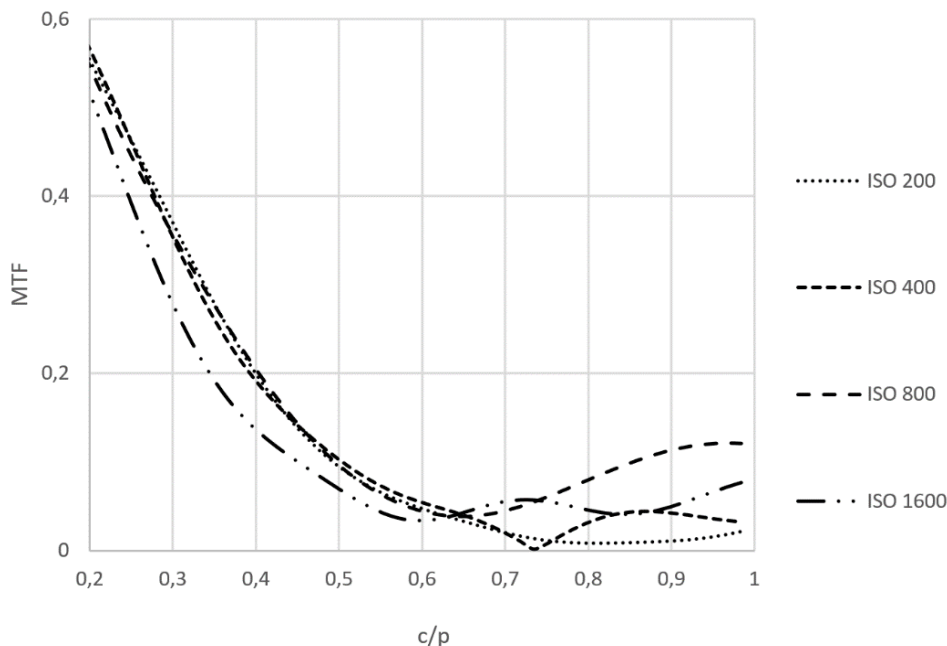


Abb. 5: MTF-Testchart: Artefakte, hervorgerufen durch hohe ISO-Einstellungen

## 2.5 Objektiv und Sensor

Die einzelnen Hersteller von UAV geben meist durch spezielle Halterungen oder durch eine feste Montage das jeweilige Kamerasystem vor. In der Praxis bleibt dem Anwender meist nur die oben beschriebenen Einstellungen am Kamerasystem optimal einzustellen, um das Auflösungsvermögen möglichst nicht zu verschlechtern.

In einzelnen Fällen ist durch die Halterung zwar das Kamerasystem (Body) gesetzt, der Anwender kann jedoch das einzusetzende Objektiv (im Rahmen gewisser Gewichtsbeschränkungen) frei wählen.

Als Kriterien für einen solchen Objektivwechsel zählt neben der Festbrennweite (Vermeidung von Zoom, Abb.4) auch die Verfügbarkeit eines manuellen Fokus, um Unschärfen im praktischen Einsatz zu vermeiden (Abb. 3 links). Das Zeiss Loxia 50 mm entspricht diesen Anforderungen. Es wurde im Labor unter optimalen Bedingungen (ISO 100, Blende f 5.6) analog zur Darstellung von DJI aus Abbildung 2 untersucht und mit dem Phase One RSM 80 mm f 5.6 an der iXM100 verglichen. Aufgrund des unterschiedlichen Sensorformates wurde für den Mittelformatsensor iXM eine längere Brennweite verwendet, um einen vergleichbaren Öffnungswinkel zu realisieren.

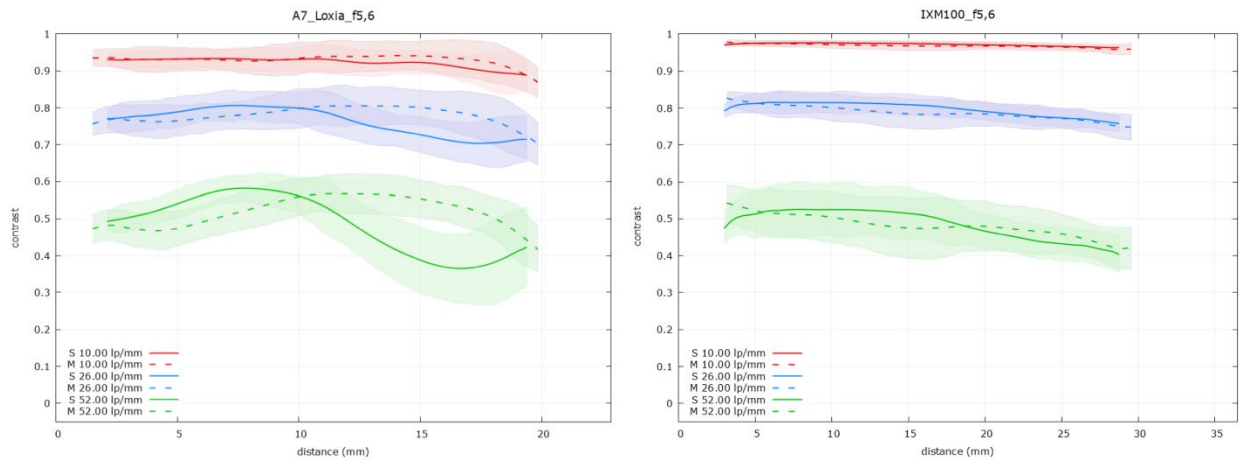


Abb. 6: Sagittale und meridional bestimmte MTF für die Sony A7 (links), rechts die Phase One iXM

Aus Abbildung 6 (links) ist ersichtlich, dass bei einem Vollformatsensor das Loxia 50 mm in der Bildmitte schlechtere Werte aufweist als das DJI DL 50 mm (Abb. 2.) Zudem weist der Unterschied zwischen sagittaler und meridionaler MTF auf einen Astigmatismus hin, der insbesondere bei feinen Strukturen zum Bildrand hin auftritt. Bessere Werte über einen, zudem größeren Bildradius, erreicht die iXM-100 von Phase One. Dieses Kamerasystem erreicht ein Auflösungsvermögen von 104 lp/mm bei MTF 30% und 156 lp/mm für eine MTF von 10%. Objektive, die vom Hersteller speziell für den Sensor gerechnet werden, weisen darüber hinaus eine geringe chromatische Aberration auf im Vergleich zu solchen Objektiven, die durch Dritthersteller gefertigt werden. Dies zeigt der Vergleich der chromatischen Aberration des Zeiss Loxia 50 mm, des Phase One RSM 80 mm und des fest mit der Sigma DP2 Merrill verbauten 30 mm Objektivs, dargestellt in der folgenden Abbildung.

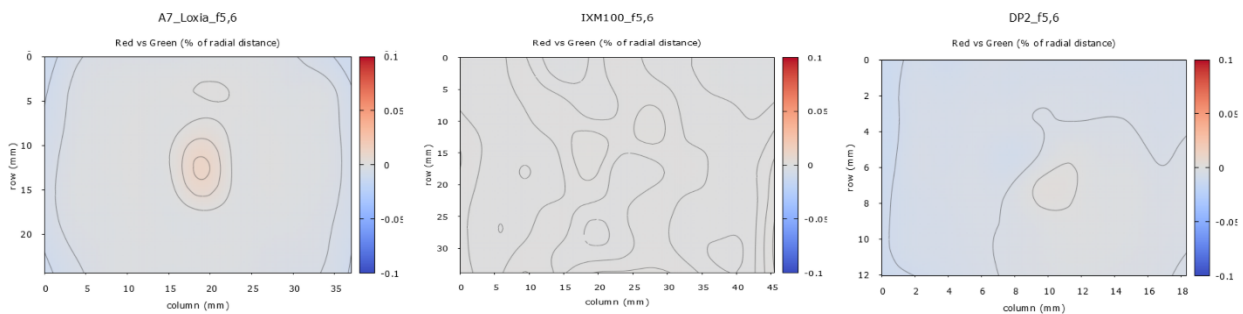


Abb. 7: Chromatische Aberration. Der Versatz roter vs. grüner Pixel ist in % des Bildradius angegeben. Links: SonyA7/Loxia, PhaseOne iXM-100 (Mitte) und Sigma DP2 Merrill (rechts).

Neben der Auswahl eines alternativen Objektivs bieten einige Hersteller ein modulares Nutzlastkonzept. Austauschbare Halterungen ermöglichen den Wechsel des Kamerasystems. Die DJI M300 kann sowohl die P1 (Vollformat) oder den Mittelformatsensor P3 tragen. Ein vergleichbares Setup wäre ein Hexapilots HP-T970 mit einer Sony Alpha 7 oder einer Phase One iXU-100.

Das Auflösungsvermögen eines digitalen Kamerasystems wird jedoch in erster Linie nicht vom Objektiv sondern vom Sensor bestimmt. Ein Beispiel hierfür sind die ausschließlich von Sigma verbauten Foveon Sensoren (GREIWE & GEHRKE 2013). Diese weisen gegenüber Sensoren mit Tiefpassfilter und Bayer-Farbfiler die höchsten MTF-Werte auf (Abb. 8).

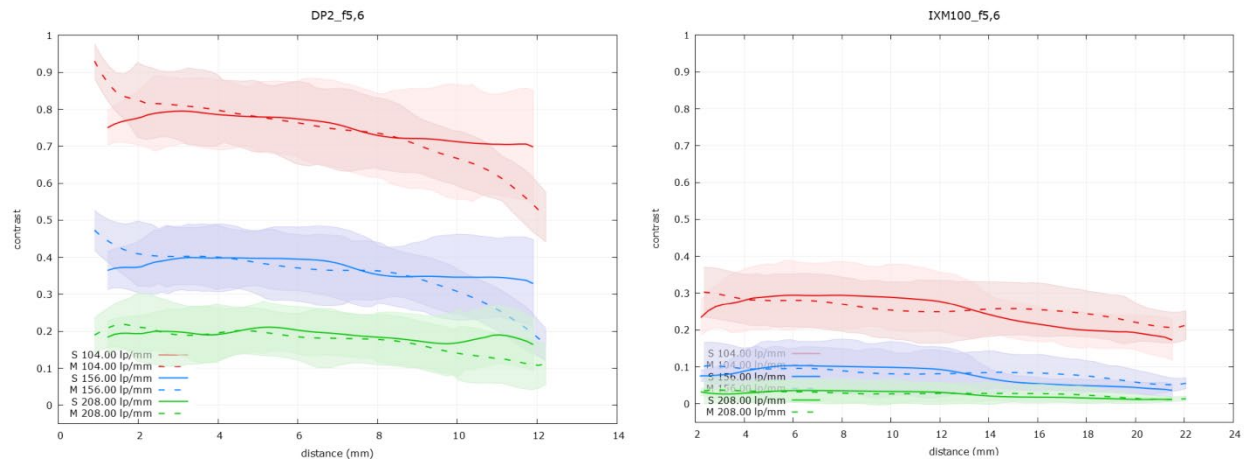


Abb. 8: MTF für Foveon-Sensor (Sigma DP2 Merrill, links) und für Bayer-Pattern (Phase One iXM, rechts).

Als Zwischenfazit kann aus den Laboruntersuchungen abgeleitet werden:

1. Der Fokus sollte vor dem Einsatz auf die Flughöhe angepasst und die Schärfentiefe berücksichtigt werden, insbesondere bei der Verwendung hochauflösender Systeme mit Normalbrennweite (ab 50 mm Kleinbild). Ein mechanischer, manueller Fokus ist immer zu bevorzugen, insbesondere wenn die Kamera nach dem Einschalten (Akkuwechsel) den Fokus erneut verstellt (z.B. Kit-Objektive Fa. Sony oder fest verbaut wie Sony RX-1).
2. Die Daten sollten im RAW-Format aufgezeichnet werden, eine Konvertierung in verlustbehaftete Formate wie JPEG sollte – wenn überhaupt – mit der Software des Herstellers erfolgen, jedoch nicht in der Kamera.
3. Objektive sollten auf ihre Abbildungsqualität hin untersucht werden. Ein einfaches Abblenden (meist von f 2.8 auf f 4.0) verbessert insbesondere am Bildrand die Bildqualität deutlich. Lediglich fest montierte, vom Hersteller auf den Sensor abgestimmte Optiken erreichen mit großer Blende ebenfalls eine gute MTF (Abb. 7).

Das Sensorsystem FOVEON wird seit fast 10 Jahren nicht mehr gefertigt. Welche Systeme heute in der Lage sind, vergleichbare Punktwolkenqualitäten zu erzeugen, ist Gegenstand im folgenden beschriebenen Praxistests am Testfeld der Hochschule Bochum.



### 3 HsBo 2021 Kampagne

#### 3.1 Zielstellung

Im Jahr 2021 war durch die digitale Lehre im sogenannten COVID-Lockdown die Möglichkeit geboten, an vielen Wochentagen<sup>3</sup> ein neu eingerichtetes Testfeld auf dem Campus der Hochschule Bochum zu befliegen. Hochschuleigene Copter und Systeme von externen Dienstleistern wurden verwendet, um eine Auswahl an Sensoren von der Sony RX-0 bis zur Phase One iXU-100 durchzuführen. Schwerpunkt der Auswertung ist jedoch nicht der Vergleich der Ergebnisse aus einer Bündelblockausgleichung. Vielmehr soll aufbauend auf die Kapitel 1 und 2 nun die Qualität der resultierenden 3D-Punktwolken hinsichtlich Auflösung, geometrischer Genauigkeit und Vollständigkeit betrachtet werden.

#### 3.2 Testfeld der Hochschule Bochum

Das Testfeld umfasst große Teile des Campus der Hochschule Bochum und hat eine Ausdehnung von ca. 230 m x 200 m. Auf dem Gelände befinden sich zwei Gebäude mit 26 und 35 m Höhe, deren Dachflächen, teilweise in unterschiedlichen Geschosshöhen, begehbar sind (Abb. 9).

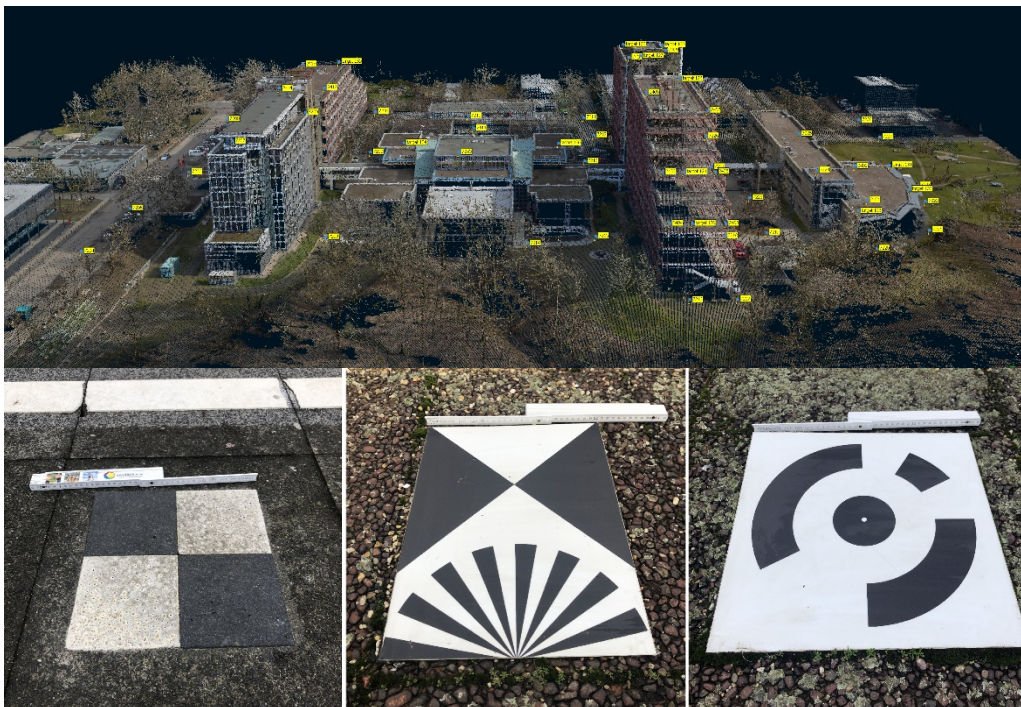


Abb. 9: Testfeld der Hochschule Bochum (oben) mit dauerhaft signalisierten Passpunkten (unten).

Neben 31 Punkten auf der Geländeoberfläche (Schachbrettmuster aus thermoplastischem Kunststoff) sind weitere 14 Punkte über Schachbrettmarken mit integriertem Siemensstern und 13 codierten Marken dauerhaft signalisiert.

<sup>3</sup> Luftfahrtrechtlich ist der Flug über „unbeteiligte Personen“ strikt untersagt. Der Einsatz von UAV für photogrammetrische Aufnahmen erfolgte in Absprache mit der zuständigen Luftaufsichtsbehörde nach den derzeit üblichen Regularien.

Die Koordinaten der Punkte wurden in einer tachymetrischen 3D-Netzmessung im November 2020 mit einer Standardabweichung von 3 mm in Lage und Höhe bestimmt. Somit stehen 56 Punkte als Pass- oder Kontrollpunkte zur Verfügung, die das Messvolumen vollständig umschließen. Testfelder, die nicht über diese räumliche Ausprägung verfügen, erlauben keine sichere Bestimmung der Inneren Orientierung im Rahmen einer in situ Kalibrierung (FRASER 2018). So kann ein Kamerasystem einerseits in einer Laborkalibrierung an einem 3D-Testfeld eine stabile Innere Orientierung aufweisen und andererseits bei einer in situ Kalibrierung auf dem Testfeld andere Werte aufweisen. Die Ergebnisse aus CRAMER et al. (2017) lassen den Schluss zu, dass auch mit einer kreuzweisen Befliegung in unterschiedlichen Höhen die Kamerakonstante nicht zuverlässig geschätzt werden kann (Abb. 10).

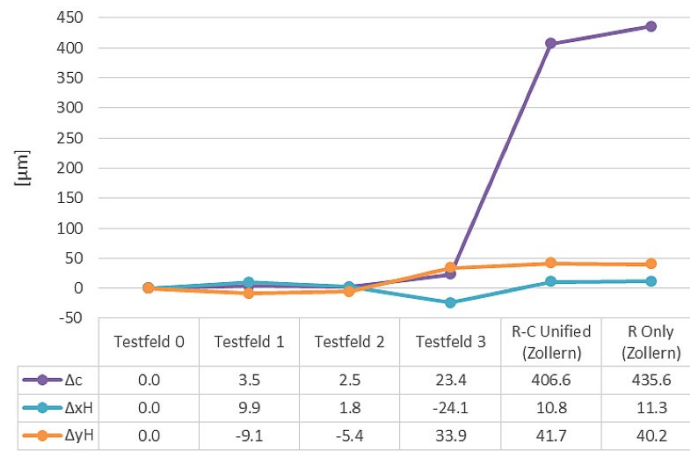


Abb. 10: Vergleich der Parameter der Inneren Orientierung zwischen Labortestfeld (3D) und in situ Kalibrierung mit Bodenpasspunkten (aus: CRAMER et al. 2017)

Diese fehlerhafte Bestimmung der Kamerakonstante durch systematische Höhenabweichungen an den Kontrollpunkten konnte am Testfeld Bochum dadurch nachgewiesen werden, da zur in situ Kalibrierung nur die 31 Punkte auf der Geländeoberfläche verwendet wurden und die 27 Punkte auf den Gebäuden als Kontrollpunkte (Abb. 11).

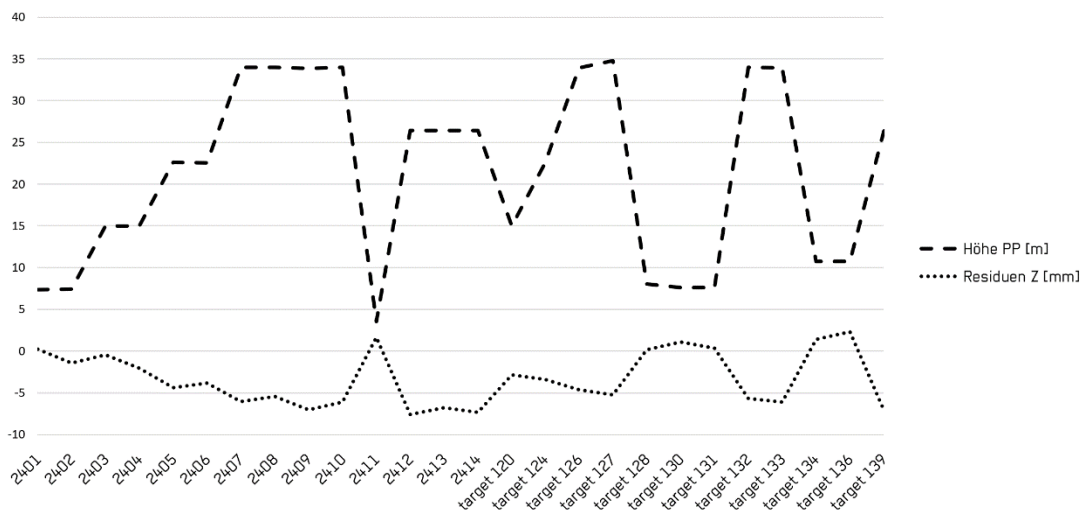


Abb. 11: Höhenabweichungen an Kontrollpunkten in Abhängigkeit von der Höhe des Kontrollpunktes

### 3.3 Kampagnen

Der Schwerpunkt der vorgestellten Arbeiten liegt jedoch nicht in der Untersuchung von Aufnahmeanordnungen und deren Auswirkung auf die Ergebnisse einer Bündelblockausgleichung. Vielmehr sollten mit möglichst vielen unterschiedlichen Sensoren, mit unterschiedlichen Einstellungen und unter Verwendung von allen Passpunkten die zugehörigen hochdichten Punktwolken erzeugt werden. Die einzelnen Systeme und Missionsparameter sind in Tabelle 1 aufgeführt.

Tab. 1: Systeme und Missionsparameter Kampagne HsBo2021

	Phantom 4 RTK SDK Edition	Sony RX-0	Sigma DP2 Merrill (DP2Mlo)	Sigma DP2 Merrill (DP2Mhi)	DJI P1	Sony Alpha 7R IV	Phase One iXM-100
Datum	16.06.2021	16.04.2021	10.06.2021	16.06.2021	10.06.2021	10.06.2021	19.03.2021
Bildmaterial	JPEG	RAW	RAW	RAW	JPEG	RAW	RAW
Belichtungsmodus	Automatik, ISO fest	manuell	manuell	manuell	Automatik	manuell	shutter priority
ISO	100	100	100	100	variabel	100	variabel
Blende	variabel	4,0	2,8	2,8	5,6	4	5,6
Belichtungszeit [1/s]	400	1000	1000	1000	1000	2000	1000
Pixel Size [ $\mu\text{m}$ ]	2,41	2,41	10,30	5,10	4,39	3,79	3,76
Brennweite [mm]	8,8	8,1	30,0	30,0	35,0	35,0	35,0
Flughöhe [m]	70	50	50	50	70	85	80
GSD [mm]	19	15	18	9	9	9	8
Fluggeschwindigkeit	5,6	4,3	3,9	1,8	1,1	4,8	4,5
Bewegungsunschärfe [pix]	0,7	0,3	0,2	0,2	0,1	0,3	0,5
RMS Bildmessung [pix]	0,17	0,87	0,18	0,17	0,15	2,73	0,13
Einbau der Kamera	quer	längs	längs	längs	quer	quer	quer
Längsüberdeckung	79	93	77	75	93	92	92
Querüberdeckung	78	78	60	60	77	79	88
Bildanzahl	188	858	488	415	698	577	635

Trotz der unterschiedlichen Sensorgrößen und Gewichten der Nutzlasten (Gimbal+Kamera) wurde durch Anpassung der Flughöhe für die kleineren Sensorformate (Phantom4, Sony Rx0) ein GSD von ca. 20 mm und für die APS-C bzw. Voll- oder Mittelformatsensoren ein GSD von ca. 10 mm geplant. Als maximale Einsatzzeit eines Systems wurden 60 Minuten über dem Testfeld zugelassen, um mehrere Systeme an einem Tag einsetzen zu können. Die Unterschiede in der Bildanzahl und damit auch der Überdeckungsparameter wird als Folge in Kauf genommen.

Die Aufnahmeeinstellungen wie ISO, Blende und Belichtungszeit wurden für die Systeme vorgegeben. Die Flüge der iXM-100 und der DJI P1 wurden von externen Projektpartnern

durchgeführt, welche eigene Kameraeinstellungen verwendet haben. So ist bei der iXM-100 die ISO freigegeben (Abb. 12), bei der DJI P1 neben der ISO zusätzlich die JPEG-Konvertierung in der Kamera aktiviert, so dass keine Rohdaten für das Post Processing vorlagen.

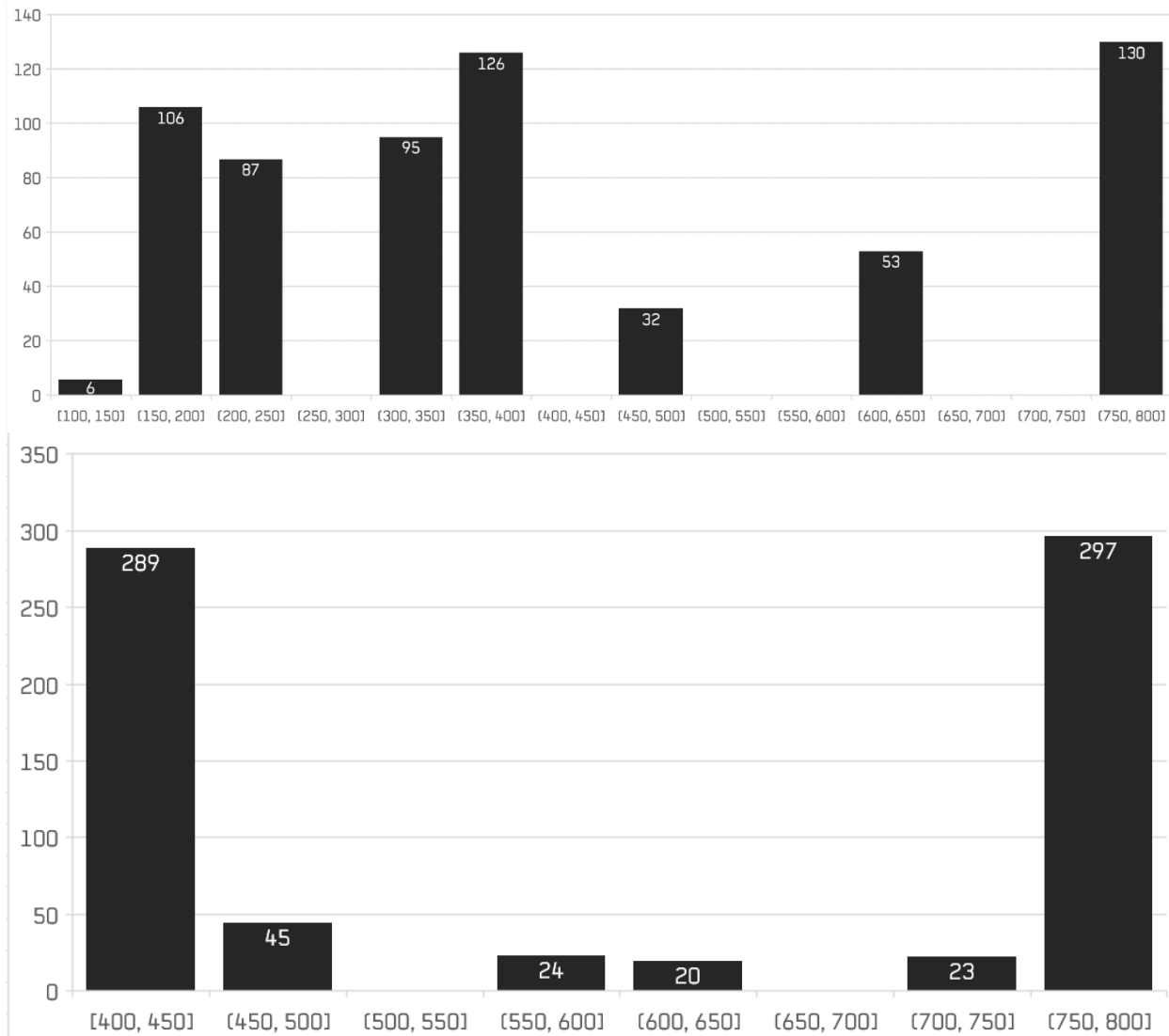


Abb. 12: Verteilung der variabel eingestellten ISO während des Fluges der iXU (oben) und der P1 (unten)

Die Punktwolken wurden in der höchsten Auflösung mit der Software Metashape erzeugt. Zur Orientierung der Bildverbände wurden alle Aufnahmen und alle vorhandenen Passpunkte verwendet, Verknüpfungspunkte nur dann zugelassen, wenn sie in drei Aufnahmen vorhanden waren. Zur Erzeugung der 3D-Punktwolke wurden Filterungen vermieden (Einstellung „mild“), im Endprodukt jene Punkte behalten, welche aus mindestens drei Tiefenbildern ermittelt wurden (Confidence-Level = 2).

## 4 Ergebnisse

Für den Vergleich der erzeugten Punktwolken wurden drei Orte auf dem Campus ausgewählt, die durch eine Detailstruktur das Auflösungsvermögen der Sensoren und die Auswirkungen der unterschiedlichen Aufnahmeeinstellungen verdeutlichen. Da die Flächen nicht mit einem Referenzsystem wie einem TLS flächig erfasst wurden, wird an dieser Stelle ein qualitativer Vergleich der Punktwolkenprofile vorgenommen.



Abb. 13: Testgebiete: "Platten" (links), "Metalldach" (Mitte) und "Zahnrad" (rechts)

### 4.1 Steinplatten

Auf einem Dach, in einer Höhe von etwa 4,5 m über dem Gelände, befinden sich in Stufen gestapelte „Waschbetonplatten“, die während der gesamten Kampagne unverändert blieben. Mit dem GSD der Sensoren von 10 bis 20 mm (Tab. 1) ist dieser Stapel mit Stufen von deutlich über 10 cm also dazu geeignet, Kanten im Oberflächenmodell nachzuweisen.

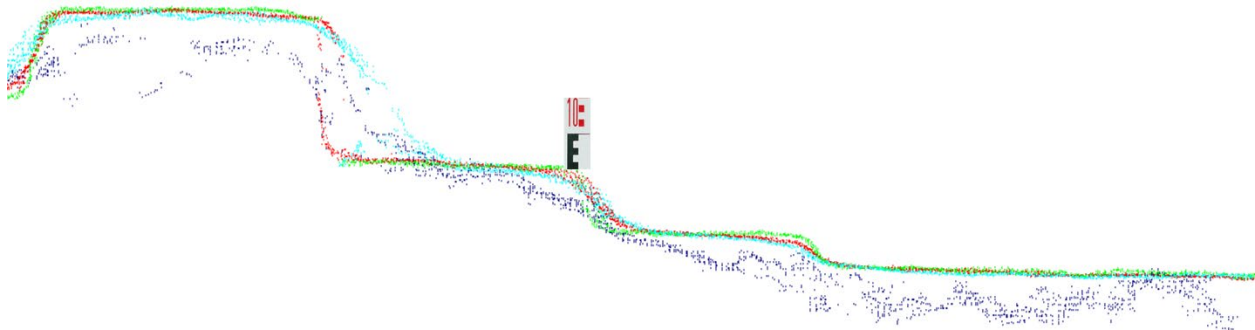


Abb. 14: Steinplatten im Profil, iXM (rot), Sony A7 (blau), DJI P1 (cyan), Sigma DP2Mhi (grün). Höhe des Maßstabs 10 cm, mit 1cm-Teilung.

In der Profildarstellung der Punktwolken (Abb. 14) weist der Datensatz der iXM aufgrund der geringen GSD die meisten Punkte und eine vollständige Erfassung der Stufen auf. Der Datensatz der Sigma DP2M weist weniger Punkte aber tendenziell schärfere Kanten auf, obwohl der Überdeckungsgrad deutlich geringer war (Tab. 1). Die Daten der DJI P1, insbesondere der Sony A7 erfassen die Stufen deutlich schlechter.

Von den Datensätzen mit dem größeren GSD (15mm bis 19 mm, Tab. 1) werden die Stufen im Vergleich zur iXM, die in Abbildung 15 als Referenz dient, zwar erfasst, allerdings wird die Kante im rechten Bereich aufgrund dieses geringen GSD nicht gut abgetastet.

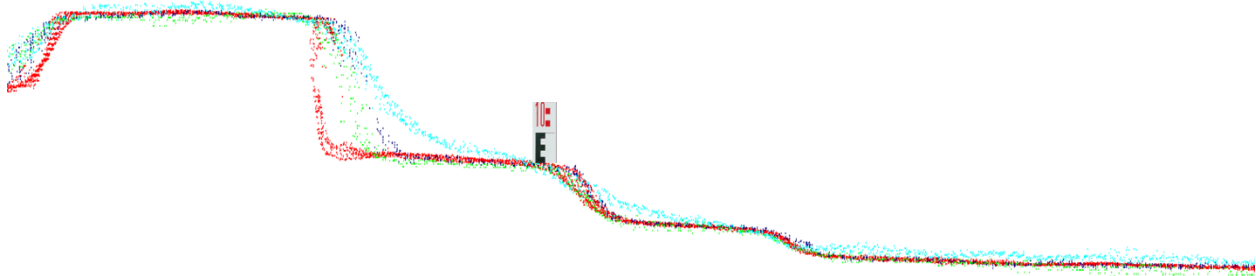


Abb. 15: Steinplatten im Profil, iXM (rot), Sigma DP2Mlo (blau), Sony RX-0 (cyan), DJI P4 (grün). Höhe des Maßstabs 10 cm, mit 1cm-Teilung.

## 4.2 Metaldach

Die Strukturen am Metaldach eines Fahrradständers sind wesentlich feiner und einhergehend mit der hellen Dachfärbung auch eine Herausforderung an die Dynamik des aufnehmenden Sensors. Hier schneidet die DJI P1 besser ab als an den Steinplatten (Abb. 14), obwohl die ISO in beiden Fällen für die entsprechenden Aufnahmen bei 800 lag. Die Sony A7 weist auch hier wiederum die schlechteste Abtastung der Oberfläche auf.

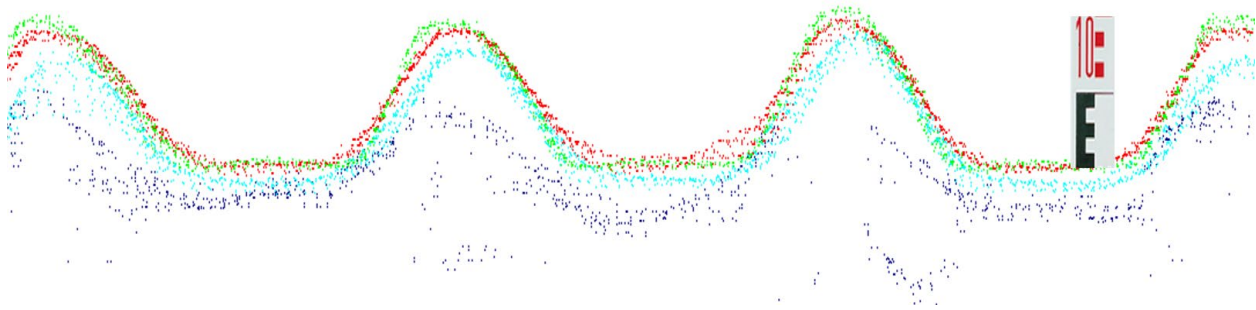


Abb. 16: Metaldach im Profil, iXM (rot), Sony A7 (blau), DJI P1 (cyan), Sigma DP2Mhi (grün). Höhe des Maßstabs 10 cm, mit 1cm-Teilung.

Im Vergleich zur Referenz (iXM, Abb. 17) weist die Sigma DP2Mlo die beste Abtastung auf, die Phantom 4 erscheint ein wenig gefiltert. Die RX-0 weist eine noch stärkere Filterung auf, die sich wie ein Höhenfehler auf die resultierende Punktwolke auswirkt.

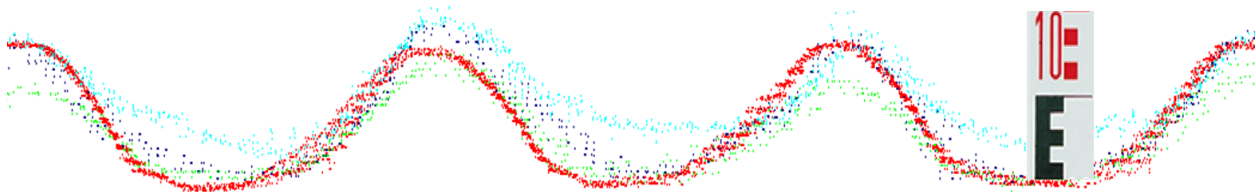


Abb. 17: Metaldach im Profil, iXM (rot), Sigma DP2Mlo (blau), Sony RX-0 (cyan), DJI P4 (grün). Höhe des Maßstabs 10 cm, mit 1cm-Teilung.

### 4.3 Zahnrad

Vor der Hochschule Bochum ist ein Industrieanlagen-Zahnrad in den Boden eingelassen, das einen Durchmesser von 7 m aufweist. Die Zahnhöhe beträgt 10 cm, die Zahnteilung ca. 20 cm. Durch den stahlgrauen Anstrich ist die Oberfläche texturarm. Eine hohe ISO-Einstellung und eine Nachschärfung von Bilddaten fällt an dieser Stelle im Untersuchungsgebiet besonders schnell auf, wie die Abbildung 18 zeigt. Die Daten der Sigma DP2 Merrill (rechts oben) werden als Rohdaten mit ISO 100 aufgezeichnet, wo hingegen die Datenaufzeichnung der DJI P1 mit variabler ISO (in diesem Fall 400) und JPEG-Konvertierung in der Kamera erfolgt, was durch die nachgeschärften Kanten in Abbildung 18 rechts unten erkennbar ist. Diese Bildartefakte führen dann zu fehlerhaften Punktwolkenprofilen.

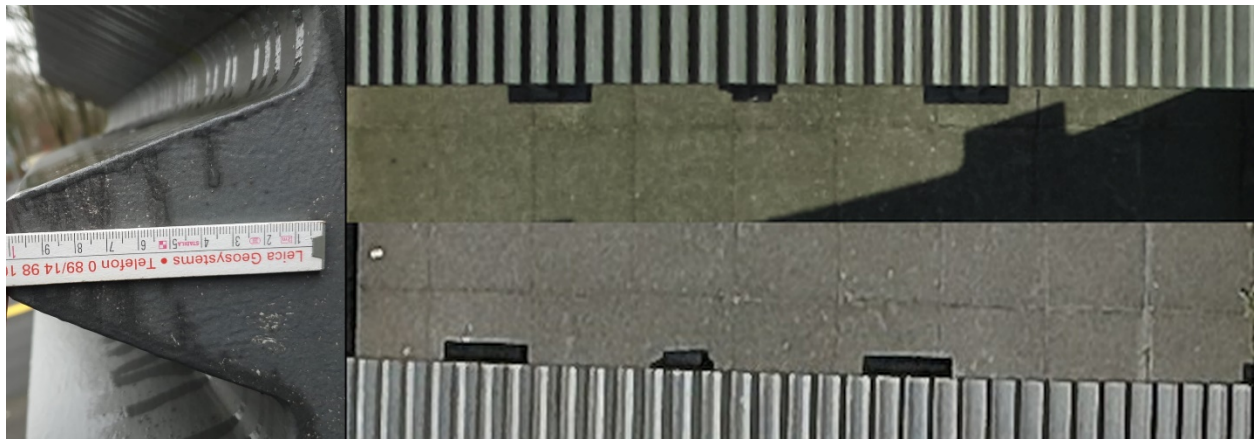


Abb. 18: Gemessene Zahnhöhe (links), Aufnahme der Sigma DP2M (oben) und der DJI P1 (unten), beide GSD ca. 10 mm.

Aus Abbildung 19 ist ersichtlich, dass selbst die iXM die Zahnhöhe von 10 cm nicht vollständig erfasst, dies gelingt lediglich der Sigma DP2 Merrill (DP2Mhi) mit gleichem GSD von ca. 10 mm. Die Daten der DJI P1 und die Sony A7 erfassen die Geometrie nicht.

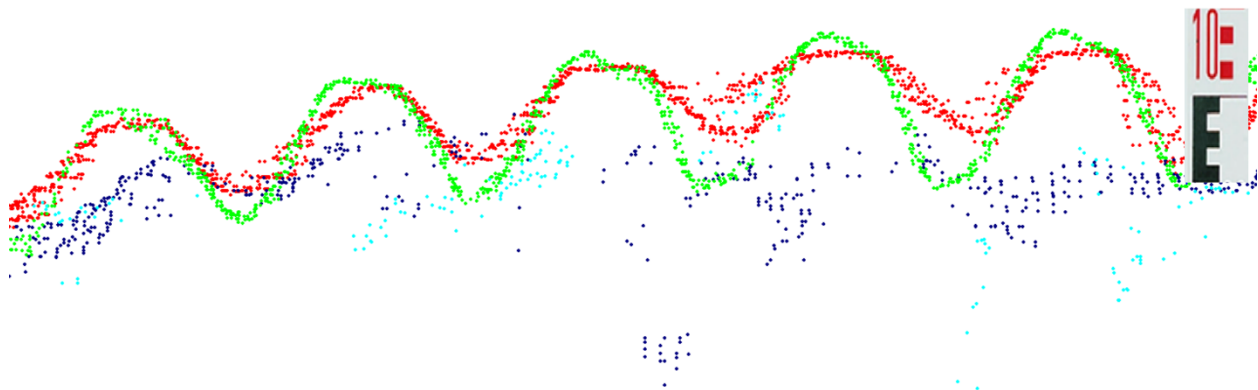


Abb. 19: Zahnhöhe im Profil, iXM (rot), Sony A7 (blau), DJI P1 (cyan), Sigma DP2Mhi (grün)

Der Datensatz der Sigma DP2 Merrill mit dem GSD von 20 mm (Mlo) zeigt im Vergleich mit der iXM (10 mm GSD) eine vollständige Erfassung des Bogens. Durch die geringe Abtastung (20 mm GSD) kann die Zahnhöhe allerdings nicht korrekt erfasst werden. Die Daten der Sony RX-0 sind hingegen stark geglättet, die DJI Phantom 4 versagt genauso wie die DJI P1 (Abb. 20, links unten in grün dargestellt). Als Grund wird hier die integrierte JPEG-Kompression in Kombination mit einer bis ISO 800 freigegebenen Aufnahmeeinstellung angenommen. Die Sony RX-0 zeichnet mit ISO 100 die Bilddaten als Rohdaten auf. Zwar weisen diese Daten auch eine Streuung auf, besitzen allerdings nicht die Artefakte der DJI Phantom 4 (Abb. 20, unten links, grün).

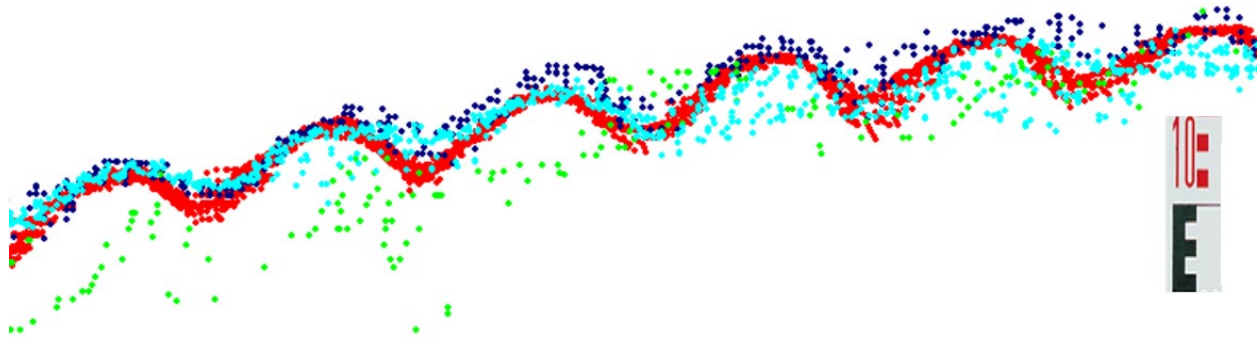


Abb. 20: Zahnhöhe im Profil, iXM (rot), Sigma DP2Mlo (blau), Sony RX-0 (cyan), DJI P4 (grün). Höhe des Maßstabs 10 cm, mit 1cm-Teilung

## 5 Fazit

Viele Untersuchungen UAV-getragener Kamerasysteme wurden auf Basis der geometrischen Betrachtung an einzelnen Kontrollpunkten im Rahmen der Bündelblockausgleichung durchgeführt. In diesem Vergleich sollten die Auswirkungen von Kameraeinstellungen und Sensor/Optik Kombination auf die Punktwolkenqualität untersucht werden.

Die enttäuschende Leistung der Sony Alpha 7R ist auf die falsche Auswahl des Objektivs zurückzuführen. Für die Portraitaufnahme optimiert, erzeugte diese Optik auch bei hohen Blendeneinstellungen keine zufriedene Bildschärfe, was sich in den Vergleichen der Testflächen bestätigte. Die Aufnahmen der DJI P1 wurden mit zu hohen ISO-Einstellungen und integrierter JPEG-Konvertierung durchgeführt, die Phantom 4 kann ohnehin nur JPEG aufzeichnen (Phantom 4 RTK, SDK Edition). Selbst bei optimalen Einstellungen kommen die kleinformatischen Sensoren an ihre Grenzen, wie die Sony RX-0 bestätigt.

Die untersuchte Phase One iXM-100 hat durch ihre hohe Bildfolgezahl und dem Sensorformat mit einer deutlich höheren Überdeckung und letztendlich auch einer intelligenteren Auswahl der freigegebenen ISO (Abb. 12) die beste Punktwolke erzeugt.

GREIWE et al. (2017) zeigten bereits in früheren Untersuchungen an einem Testfeld in Dortmund (Zeche Gneisenau), dass die Sensortechnik des Foveon-Sensors im Vergleich zu den CMOS-Kameras mit Bayer-Pattern aufgrund der besseren MTF schlussendlich auch die bessere geometrische Genauigkeit und Oberflächenabtastung ermöglicht.

Allerdings wird der Foveon-Sensor exklusiv von SIGMA verbaut und hat sich seit Produktionsende 2013 im Markt nicht durchgesetzt. Sie wird vermutlich auch nicht weiter



produziert werden. Es ist zudem zu sehen, dass sich CMOS-Sensoren mit Bayer-Pattern am Markt durchsetzen. Somit muss bei der Ableitung von hochgenauen und hochauflösenden Oberflächenmodellen aus solchen Luftbildern zukünftig ein Augenmerk auf den Praxiseinsatz mit Bayer-Pattern gelegt werden. Hier wird aus Sicht der Autoren in Zukunft zu prüfen sein, inwiefern sich die De-Bayering / De-Mosaicking und Binning – Methoden für CMOS-Sensoren insoweit verbessern lassen können, bis sie sich mit den herkömmlichen CCD-Sensoren messen können.

Es wird den Herstellern, aber auch den Dienstleistern für UAV-Befliegungen empfohlen, Praxistests mit ihren Systemen auf solchen UAV-Testfeldern durchzuführen, welche die vom Anwender aufzunehmenden Objekte in Ausdehnung und Höhenstaffelung repräsentieren, um hiermit deren Eignung für eine verlässliche Aufnahme des Objektes zu gewährleisten.

## 6 Literaturverzeichnis

- CRAMER, M., PRZYBILLA, H.-J., MEIBNER, H. & STEBNER, K., 2017: Kalibrierung und Qualitätsuntersuchungen UAV-basierter Kamerasysteme. Unmanned Aerial Vehicles 2017 (UAV 2017). DVW-Schriftenreihe, **86**, DVW e. V. (Hrsg.), 67-84.
- CUNNINGHAM I. A. & FENSTER, A. 1987: A method for modulation transfer function determination from edge profiles with correction for finite-element differentiation. Medical Physics, 14(4), <https://doi.org/10.1118/1.596064>.
- DJI, 2021: Zenmuse P1 User Manual, V 1.2 (04.2021), <http://www.dji.com/zenmuse-p1>
- BOREMANN, G., D., 2001: Modulation Transfer Function in Optical and Electro-Optical Systems. SPIE – The International Society for Optivcal Engineering, Bellingham, Washington
- GHAASSOUN, Y., GERKE, M., KHEDAR, Y., BACKHAUS, J., BOBBE, M., MEISSNER, H., TIWARY, P.K. & HEYEN, R., 2021: Implementation and Validation of a High Accuracy UAV-Photogrammetry Based Rail Track Inspection System. Remote Sens., **13**, 384, <https://doi.org/10.3390/rs13030384>.
- FRASER, C., 2018: Camera Calibration Considerations for UAV Photogrammetry. Towards Photogrammetry 2020, ISPRS Technical Comm. II Symposium 2018, Riva del Garda, Italy, 3-7 June 2018.
- GREIWE, A. & GEHRKE, R., 2013: Foveon Chip oder Bayer Pattern – geeignete Sensoren zur Aerophotogrammetrie. Photogrammetrie, Laserscanning, Optische 3D-Messtechnik - Beiträge der Oldenburger 3D-Tage 2013, Luhmann, T. & Müller, C. (Hrsg.), VDE Verlag GmbH, Berlin/Offenbach, 334-343.
- GREIWE, A., PETERS, J., SPRECKELS, V., SCHLIENKAMP, A., LÖSLER, M. & ESCHELBACH, C., 2017: Geeignete Kameras zur Geländeaufnahme durch UAS. Photogrammetrie, Laserscanning, Optische 3D-Messtechnik - Beiträge der Oldenburger 3D-Tage 2017, Luhmann, T. & Müller, C. (Hrsg.), VDE Verlag GmbH, Berlin/Offenbach, 202-221.
- ISO 12233:2017(E), 2017: Photography – Electronic still picture imaging – Resolution and spatial frequency responses, Standard. International Organization for Standardization, Genf.
- LUHMANN, T., 2018: Nahbereichsphotogrammetrie Grundlagen - Methoden - Beispiele. 4. Auflage, Wichmann Verlag, Wiesbaden

- MEIßNER, H. PRZYBILLA, H.-J., CRAMER, M. & STEBNER, K., 2018: UAV-Kameras – Bewertung von Kalibrierung, Langzeitstabilität und optischem Auflösungsvermögen. Unmanned Aerial Vehicles 2018 (UAV 2018). DVW-Schriftenreihe, **89**, DVW e. V. (Hrsg.), 47-59.
- NASSE, H.H., 2008: Wie liest man MTF-Kurven? Technischer Bericht, Geschäftsbereich Photo Objektive, Carl Zeiss
- REICHENBACH, S. E., PARK, S. K. & NARAYANSWAMY, R., 1991: Characterizing digital image acquisition devices. Optical Engineering, **30**, 170-178.
- SPRECKELS, V., SCHLIENKAMP, A., GREIWE, A. & EBERHARDT, L., 2016: Eignung von ALS, Aero- und UAS-Photogrammetrie zur Früherkennung und Erfassung von Tagesbrüchen. Publikationen der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation e.V., **25**, 97-112.
- PRZYBILLA, H.-J. & KERSTEN, T., 2022: Aspekte der Qualitätssicherung für UAV-Anwendungen in der Photogrammetrie. Qualitätssicherung geodätischer Mess- und Auswerteverfahren 2022, Beiträge zum 208. DVW-Seminar am 2. und 3. Juni 2022 in Berlin, 103-121.
- PRZYBILLA, H.-J., 2022: Wer wird denn gleich in die Luft gehen? UAV 2022 – Innovation und Praxis. DVW-Schriftenreihe, **100**, DVW e. V. (Hrsg.), 17-38.
- VERHOEVEN, G. J. J., 2010: It's all about the format – unleashing the power of RAW aerial photography. International Journal of Remote Sensing, **31**(8), 2009-2042.
- ZURHORST, A. & HUDASCH, J., 2020: Baubegleitende Dokumentation mittels UAV und bildbasierten Punktwolken. Unmanned Aerial Vehicles 2020 – The Sky is the limit? DVW-Schriftenreihe, **97**, DVW e. V. (Hrsg.), 219-228.

# Integration von 3D-Laserscanning in die Rohstoffgewinnung im Untertagebau

JÜRGEN SELENSKI<sup>1</sup>, OLIVER REUß<sup>1</sup>, JONAS ROTH<sup>1</sup>, VERENA HÄNDLER<sup>1</sup>, TOMASZ BADECKI<sup>2</sup>, UWE SCHIRMER<sup>2</sup>, MARKUS BOGNER<sup>3</sup>, ANDREAS WAGNER<sup>3</sup> & ANSGAR BRUNN<sup>1</sup>

*Zusammenfassung: Im regelmäßigen Untertagebau wachsen beständig die Anforderungen an Wirtschaftlichkeit und Dokumentation. Eine detaillierte Situationserfassung durch terrestrisches Laserscanning kann die Datengrundlagen schaffen. Die Vermessungsaufgaben fallen dabei sporadisch regelmäßig an und sind im Regelbetrieb durch Nichtvermesser durchzuführen. In diesem Beitrag werden Ansätze zur Vorbereitung eines vermessungstechnischen Umfelds entwickelt, die die Bedürfnisse des Untertagebaus erfüllen. Dazu werden die Aufgaben der Volumenbestimmung des Abbaus, die Optimierung der Bohrlochtiefen und die Dokumentation des Gesamtabbaus exemplarisch behandelt. Die vorgestellten Verfahrensabläufe wurden exemplarisch realisiert. Die Ergebnisse zeigen, dass die Verfahren die erwartete und notwendige Vollständigkeit sowie auch Genauigkeit liefern.*

## 1 Einleitung

In Industrie, Bauwesen und Handwerk wird das Potenzial von dichten 3D-Daten zunehmend genutzt. In der Regel werden dabei digitale Teilschritte in bestehende Arbeitsprozesse integriert. Andauerndes Ziel ist es, repetitive Arbeitsabläufe und Prozessketten (teilweise) zu automatisieren und damit wirtschaftlicher zu gestalten sowie zusätzliche Informationen zu gewinnen. Solche wiederkehrenden Arbeitsabläufe sind auch bei der Gesteinsförderung im Untertagebergbau zu finden. Die Firma Knauf Gips KG baut in der Grube Hüttenheim IV im Örtferfestenbau Anhydrit ab. Dabei entsteht ein sich wiederholendes System aus Festen (Pfeilern), die stehen bleiben, und Kammern (Örtern), aus denen der Anhydrit gewonnen wird (Abb. 1).

### 1.1 Anforderungen des Untertagebaus

Für die Gewinnung des Anhydrits wird ein konventionelles Bohr – und Sprengverfahren eingesetzt. Hierzu werden Örter (Hohlräume, Kammern) mit einer Breite von 7 m und einer Höhe von 4 m aufgefahren, sowie schachbrettmusterartig Festen (Pillar; Abb. 1) mit den Abmessungen 10 m lang und 6 m breit stehen gelassen.

Der Gewinnungszyklus untertage setzt sich wie folgt zusammen: 1. Bohren der Sprenglöcher, 2. Besetzen und Sprengen, 3. Beladen der Fahrzeuge, 4. Transport zum Brecher und 5. Brechen und Zerkleinern.

---

<sup>1</sup> Technische Hochschule Würzburg-Schweinfurt (THWS), Fakultät für Kunststofftechnik und Vermessung, Studienbereich Geo, Röntgenring 8, D-97070 Würzburg,  
E-Mail: [Ansgar.Brunn, Oliver.Reuss]@thws.de

<sup>2</sup> Knauf Gips KG, Willanzheim-Hüttenheim, Hüttenheim 45, D-97348 Willanzheim-Hüttenheim,  
E-Mail: [Uwe.Schirmer, Tomasz.Badecki]@knauf.com

<sup>3</sup> ANGERMEIER INGENIEURE GmbH, i\_PARK Klingholz 16, 97232 Giebelstadt,  
E-Mail: [Bogner.Markus, Wagner.Andreas]@angermeier.de

Zu 1.: In regelmäßigen Abständen werden durch den Markscheider Vermessungspunkte (Stunden) an der Firste (Decke) der Grube eingerichtet. Sie dienen dem Bohrmeister zur Ausrichtung seines Bohrwagens. Die Einrichtung erfolgt rein optisch. In zwei Stunden werden Lote eingehängt. Anschließend wird ein Strahler an das hintere Lot gehalten. Anhand der Schatten markiert ein zweiter Mitarbeiter die Mitte des Ansatzpunktes an der Ortsbrust (Abbauwand). Danach wird der Bohrwagen mittels zwei Abstandslasern positioniert, bevor das programmierte Bohrschema halbautomatisch abgebohrt wird. Es werden 4 bis 5 Abbauorte pro Tag gebohrt.

Zu 2. bis 5.: Die Bohrlöcher werden mit Sprengstoff gefüllt und anschließend gezündet. Das gesprengte Haufwerk wird mit Radladern auf die Transportfahrzeuge aufgeladen. Die Fahrzeuge transportieren das Material zum untertägigen Brecher, der den Anhydrit vorzerkleinert, bevor er zur weiteren Verarbeitung mit einer Förderbandanlage in die Fabrik gefördert wird.

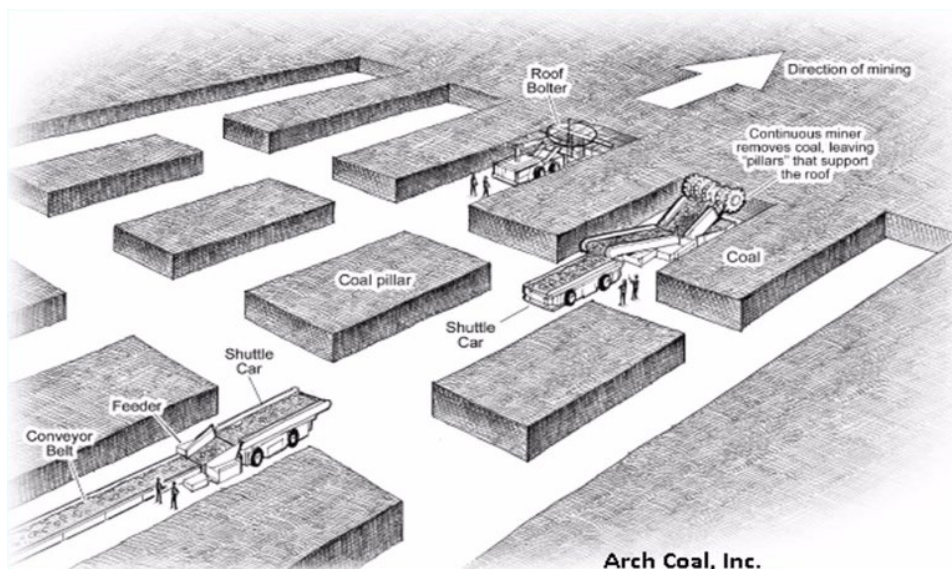


Abb. 1: Prinzip des Orterfestenbaus in der Grube Hüttenheim IV (Quelle: ARCH COAL 2023)

Die zukünftige Nutzung digitaler Mess- und Auswerteverfahren soll die Bewirtschaftung der Lagerstätte effizienter gestalten. Hierfür wurden in enger Abstimmung mit dem Bergbauunternehmen Ansätze ausgearbeitet, die das Potenzial von terrestrischem Laserscanning im Untertagebergbau aufzeigen. Wichtig war es etablierte Prozessketten zu berücksichtigen und die neuen Verfahren in die bisherigen Abläufe zu integrieren. Aus den vom Unternehmen geäußerten Zielen und der ersten testorientierten 3D-Datenerfassung stellten sich die Volumenberechnung der abgebauten Kammer pro Sprengung, die Dokumentation des Abbaus und die Sicherstellung einer einheitlichen Bohrtiefe für den Sprengvortrieb als erfolgversprechend heraus. Die Integration des terrestrischen Laserscannings soll die effektive Nutzung der Lagerstätte gewährleisten, den Bohr- und Sprengzyklus optimieren sowie den Vortrieb dokumentieren.

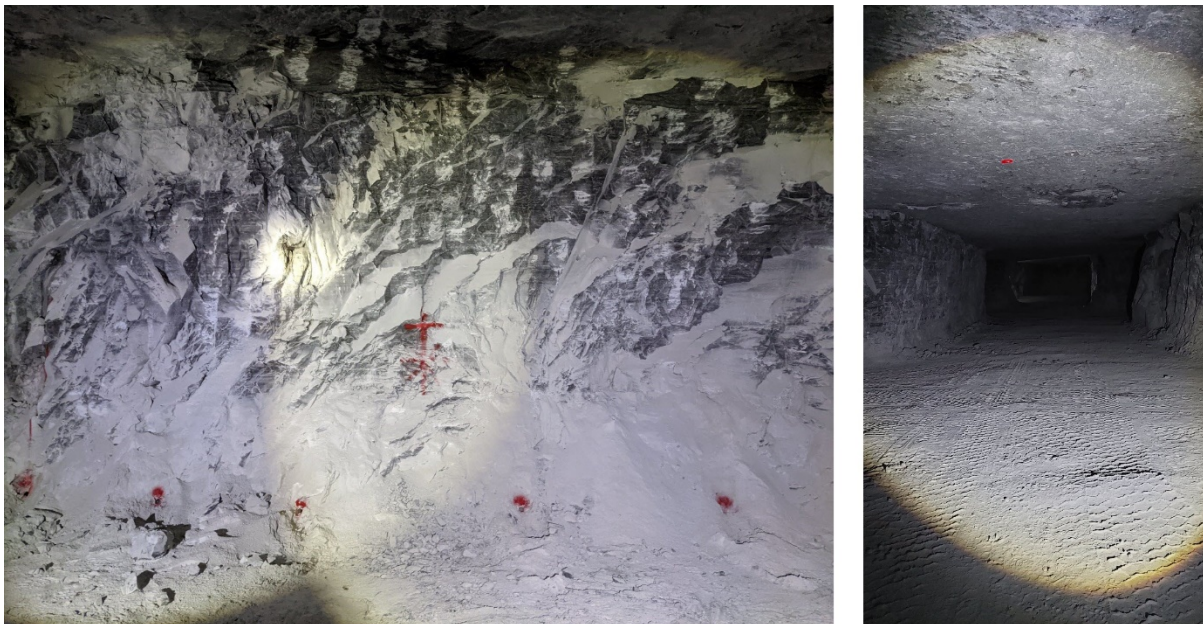


Abb. 2: Abbauwand mit Markierungen (links) und Kammersituation (rechts) untertage. Rote Markierung an der Firste im rechten Bild zeigt einen Stundenpunkt des Markscheiders

## 1.2 Vorarbeiten

### 1.2.1 Terrestrisches Laserscanning im Bergbau

Das terrestrische Laserscanning (TLS) wird seit vielen Jahren regelmäßig im Bergbau eingesetzt. Das TLS ermöglicht eine vollständige 3D Datenerfassung, ist hochgenau und auch in Kombination mit Kameras möglich, erzeugt jedoch große Datenmengen und erfordert eine schrittweise Datenerfassung (TSCHARF 2018). Der übliche Einsatzrahmen von TLS unterscheidet sich jedoch von der hier vorliegenden Situation, durch die Kombination eines regelmäßigen Untertagebaus mit dem sprengenden Abbauverfahren im Anhydritabbau und dem Wunsch nach einer Low-Cost-Lösung. Relevante aktuelle Verfahren zum terrestrischen Laserscanning innerhalb der Gewinnung von festen Stoffen lassen sich nach dem Ort des Abbaus einteilen:

Im Tagebau, z.B. in Steinbrüchen, werden terrestrische Laserscanner zur Dokumentation (z.B. OPARIN et al. 2007) und Analyse der Abbruchwand (z.B. LAUX 2017) eingesetzt, oft als Einzelstandpunkte bei überschaubaren Objektgrößen. Im großflächigen Tagebau (z.B. Braunkohle in Deutschland) werden die Daten des terrestrischen Laserscanners durch Airborne Laserscanning (JAWECKI et al. 2017) und Drohnenbefliegungen (UAV) ergänzt. In Steinbrüchen und im Tagebau allgemein wird mit Schaufelrad- oder Löffelbaggern gearbeitet. Ebenso sind Sprengungen üblich.

Im Untertage, z.B. für den Stollenabbau, hat die DMT früher bereits einen speziellen explosionsgeschützten Laserscanner mit der Zoller+Fröhlich GmbH entwickelt (KUCHENBECKER 2010), der der Richtlinie 94/9/EG (abgelöst durch die Richtlinie 2014/34/EG in 2016) sowie der BetriebssicherheitsVO der DMT genügte. STUDNICKA & GROISS (2019) berichten von einer Gesamtaufnahme eines Untertage-Bergwerks. Im europäischen Eisenbahn- und Straßentunnelbau ist die Ortsbrust i.d.R. nicht beobachtbar, da die Tunnelbohrmaschine (TBM) die gesamte Ortsbrust

durchgehend bearbeitet und verdeckt. TLS dient in diesen Fällen zur Dokumentation und Qualitätssicherung hinter der TBM bzw. der wandernden Ortsbrust, der Dokumentation von Sekundärauswirkungen der Baumaßnahme (MATWIJ et al. 2021) und auch zur Sicherstellung von Durchfahrtsprofilen im Betrieb (PEJIC 2013).

Terrestrische Laserscanner werden zunehmend durch fahrzeugbasierte Mobile Mapping Systeme dort ersetzt, wo die Mobile Mapping Systeme wirtschaftlich eingesetzt werden können. Im Tagebau (Braunkohle) werden regelmäßig wiederholend große Flächen befahren, sodass das fahrzeugbasierte Mobile aufgrund der reduzierten Auswertezeit interessant ist. Das INS und GNSS liefern im freien Gelände präzise Positionsdaten. Mit diesen Sekundärdaten lassen sich schnell Punktwolken und Oberflächenvermaschungen berechnen, die dann zur Kartierung und Volumenermittlung in Ergänzung zu UAV-Vermessungen herangezogen werden.

Die Situation im Anhydritabbau in der hier relevanten Grube Hüttenheim IV unterscheidet sich dadurch, dass hier nur kleine Bereiche jeweils im Fokus stehen, untertage gesprengt wird und regelmäßig die Oberfläche der Ortsbrust zur Optimierung weiterer Sprengungen beobachtet werden soll. GNSS scheidet als Positionierungssystem untertage aus. Zentimetergenaue Einmessungen wie bei KAJZAR et al. (2015) sind nicht erforderlich. Wirtschaftliche Speziallösungen sind daher hier für den Einsatz im nichtvermessungstechnischen Regelbetrieb erforderlich. Vergleichbare Abbausituationen sind in Deutschland im Kalibergbau gegeben.

### 1.2.2 Registrierung und Georeferenzierung von 3D Punktwolken aus terrestrischem Laserscanning

Terrestrisches Laserscanning ist ein Verfahren zur Messung von Millionen von Oberflächenpunkten der vom jeweiligen Laserscannerstandpunkt aus sichtbaren Objekte bzw. Objektteile. Zur vollständigen Objekterfassung sind im Allgemeinen mehrere Standpunkte des Laserscanners erforderlich, die dann zueinander registriert und ggfls. in ein übergeordnetes Referenzsystem georeferenziert werden müssen. Die Georeferenzierung kann nach zwei unterschiedlichen Ansätzen erfolgen:

- In einer Einstufenlösung wird jeder Laserscannerstandpunkt in das übergeordnete Koordinatensystem georeferenziert. Es werden häufig Targets als Passpunkte genutzt (z.B. DEGEN et al. 2014).
- In der Zweistufenlösung erfolgt zunächst eine gegenseitige Registrierung der einzelnen lokalen Punktwolken anhand homologer Messungen (z.B. von Targets (Target-to-Target Registrierung), automatisch detektierter Ebenen (Plane-to-Plane Registrierung) oder lokaler Punktgeometrien (Cloud-to-Cloud Registrierung). Dieser Vorgang kann als relative Orientierung ausgefasst werden. In einem zweiten Schritt wird die relativ orientierte Punktwolke in das übergeordnete Referenzsystem georeferenziert (Absolute Orientierung).

Die Zweistufenlösung erfordert mehr Scanaufwand, weniger Passpunkte und hat eine homogenere innere Geometrie, weil weniger äußere Einflüsse auf die Punktwolkengeometrie wirken. Sequentielle Anschlüsse von Punktwolken an vorhandene Registrierungen verhalten sich ungünstig in Bezug auf die absolute (äußere) Genauigkeit (WUJANZ et al. 2019).

## 2 Untersuchungen

Die Untersuchungen umfassen drei Aufgabenbereiche: a) Ermittlung des Volumens der Abbruchkammer, b) Optimierung der Bohrlochtiefen und c) die Dokumentation des Gesamtabbaus. Aufgrund der Einfachheit der Bedienung und der erprobten Genauigkeit kommen terrestrische Laserscanner zum Einsatz. Handgetragene Scanlösungen (wie bei HÖBELBARTH et al. 2018) schieden aufgrund der erwarteten geringen Positionierungsgenauigkeit in dieser regelmäßigen Abbausituation als Messsensoren aus. Die 3D-Vermessung mittels terrestrischem Laserscanning kann in den genannten drei Themenbereichen die Wirtschaftlichkeit und Genauigkeit steigern. Im Einzelnen wurden drei Verfahren untersucht, die in den folgenden Abschnitten beschrieben werden.

### 2.1 Volumenberechnung der Abbaukammer

Für die Ermittlung des gesprengten Haufwerks sind zwei Laserscans erforderlich. Die Ortsbrust wird einmal vor und einmal nach der Sprengung erfasst. Die beiden multitemporalen Scans bzw. Scangruppen werden zueinander orientiert, um anschließend das geförderte Volumen ableiten zu können. Es wurden zwei Ansätze der Registrierung (Abs. 1.2.2) verfolgt:

1. Lokaler Cloud-to-Cloud-basierter Ansatz
2. Target-basierter Ansatz mit Festpunktfeld

Im Folgenden wird die Umsetzung beider Ansätze beschrieben.

#### 2.1.1 Volumenbestimmung durch einen lokalen Ansatz

In dem lokalen Ansatz erfolgt lediglich eine gegenseitige Registrierung der multitemporalen 3D-Punktwolken. Ein Bezug zu einem übergeordneten Referenzsystem wird nicht hergestellt. Nach dem Scanning der Ortsbrust und der Registrierung von ggfls. mehreren Scanstandpunkten pro Zeitpunkt werden die Bereiche in den beiden registrierten Punktwolken im Rückraum identifiziert, die sich durch die Sprengung nicht verändert haben. Die Abb. 3 (links) ist also multitemporal aufzufassen. Die Stunden (Festpunkte) sind die Vermarkungen des bisherigen Referenzsystems der Markscheider. In einer Minimalkonfiguration ist Standpunkt 1 der Standpunkt des Laserscanners vor der Sprengung und Standpunkt 2 der Standpunkt nach der Sprengung und dem Abtransport. Gelbe Linien markieren potentielle gleichbleibende Teile der Punktwolke, die für die Registrierung geeignet sind. Diese unveränderten Bereiche werden durch eine Cloud-to-Cloud-Zuordnung registriert. Das Ergebnis ist in Abb. 3 (rechts) dargestellt. In der Abbildung ist bei Standpunkt 9 die alte Ortsbrust zu sehen. Anmerkung: In dieser Messkampagne wurden zusätzlich Standpunkte zur Verbesserung der Prozessdokumentation gemessen.

Zum Einsatz kam der Z+F IMAGER 5016 der THWS. Die Cloud-to-Cloud-Registrierung erfolgte in der Software ZF Laser Control mit einer Genauigkeit besser als 1cm Standardabweichung. Die Volumenberechnung erfolgte in der Software CloudCompare mit dem Tool "Volumen Calculation". In diesem Fall wurden 68m<sup>3</sup> Material abgebaut. Tests ergaben, dass eine Scanauflösung "Medium" für die fachlichen Anforderungen ausreichend ist.

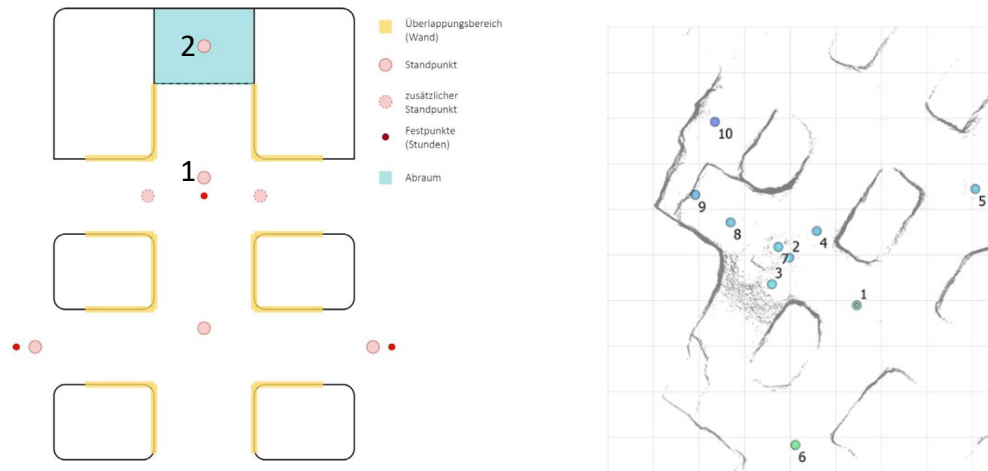


Abb. 3: Messanordnung für die lokale Volumenbestimmung des Abraums und Ergebnis der multitemporalen Messaufnahme

### 2.1.2 Volumenbestimmung mit Festpunktfeld und target-basierter Registrierung

In dem zweiten Ansatz der Punktwolkenregistrierung wurden künstliche Passpunkte – in diesem Fall Referenzkugeln – verwendet, um die Transformationsparameter zwischen den einzelnen Laserscans zu bestimmen. Vorteile dieses Verfahrens sind:

- Die Passpunkte können tachymetrisch, d. h. mit höherer Genauigkeit in einem übergeordneten Koordinatensystem bestimmt werden.
- Eine einfache Qualitätssicherung der Registrierung ist möglich, da die Residuen in den Passpunkten einfach zu deuten sind und Schwellwerte leicht zu definieren sind.
- Es sind keine Näherungswerte notwendig, was insb. im untertägigen regelmäßigen Abbauraster schwierig ist. Die Gefahr einen cloud-to-cloud-basierten Ansatz in einem falschen Angriff zu initialisieren, was in der Folge in einem falschen Ergebnis konvergieren könnte, ist hoch.

Eine weitere Nutzung des Passpunktfelds, die nicht direkt mit der Registrierung zu tun hat, ist dessen Nutzung für eine Maschinensteuerung des Bohrgeräts. Nach Ermittlung des optimalen Bohrlochrasters und -tiefen aus den Scandaten, vgl. Kap. 2.2, kann die Information digital an das Bohrgerät übergeben und z.B. für eine tachymetrische Gerätesteuerung genutzt werden.

Schwierigkeiten in der Umsetzung der target-basierten Methode sind hingegen die Sicherstellung der Stabilität der Festpunkte und die Sichtbarkeit einer ausreichenden Anzahl in den Scans. Eine sinnvolle Konfiguration, d. h. eine gute dreidimensionale Verteilung der Passpunkte muss gegeben sein.

Besonderes Augenmerk lag auf der stabilen Anbringung der Passpunkte in dem spröden Gestein, um auch den Erschütterungen in unmittelbarer Nähe Stand zu halten. Eingesetzt wurden Steckzapfenadapter, auf die sowohl ein Prisma als auch eine Scanningkugel aufgesetzt werden. Die Kombination ist so gewählt, dass das Zentrum der Kugel mit dem des Prismas übereinstimmt und keine rechnerische Korrektur notwendig ist.



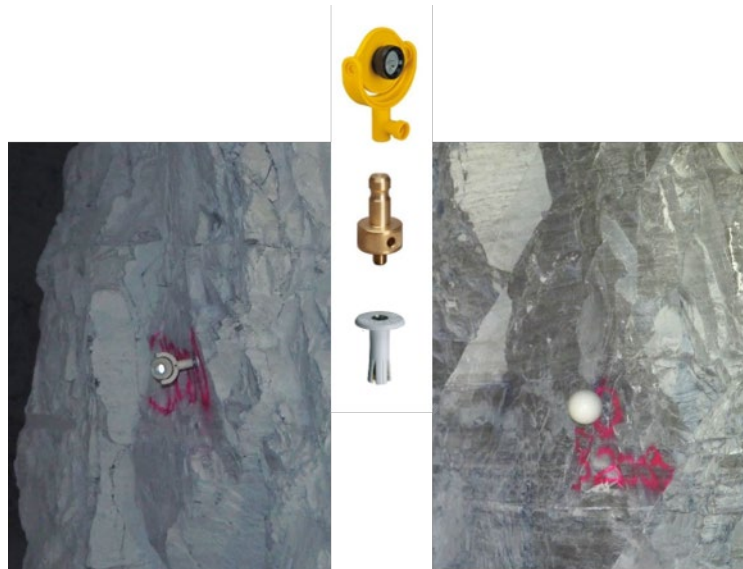


Abb. 4: Anbringen von Gewindeaufnahmen (Dübelbolzen) für Steckzapfen

Die (Scan-)Passpunkte wurden tachymetrisch im übergeordneten Hauptzugnetz des Markscheiders (Stundenpunkte wie in Abb. 2 rechts) eingemessen, so dass sich eine Einzelpunktgenauigkeit im Bereich 1-2 mm im Abbildungssystem ergibt. Die Auswertung der Scanstandpunkte erfolgt softwarebedingt in einem lokalen System, die Auswirkungen auf die zu ermittelnden Volumina sind aber vernachlässigbar.

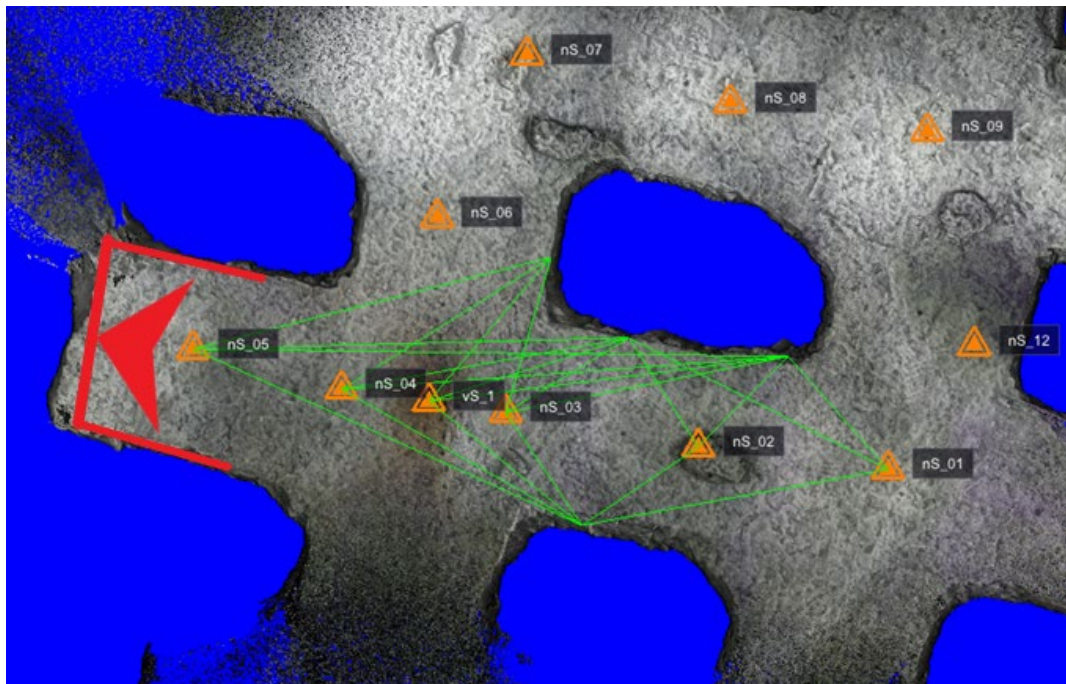


Abb. 5: Übersicht Messbereich (rot: Sprengvortrieb / orange Dreiecke: Standpunkte / grün: Passpunkt-Beziehungen)

Für die untertägige Aufnahme wurden zwei unterschiedliche terrestrische Laserscanner verwendet und deren Ergebnisse und Eignung miteinander verglichen. Für die Aufnahmen vor der Sprengung kam ein Trimble X7 zum Einsatz, nach Sprengung ein Leica RTC360. In Abb. 5 ist die Messkonfiguration für die Erfassung des rot markierten Abbaubereich als Draufsicht dargestellt. Die orangenen Dreiecke stellen die einzelnen Scanstandpunkte dar, die jeweiligen Präfixe zeigen den Messzeitpunkt vor (vS) oder nach (nS) der Sprengung. Die grünen Linien zeigen die Visuren zu den jeweils verwendeten Passpunkten. Zur besseren Interpretierbarkeit der Punktwolke, z. B. in Bezug auf die Gesteinsschichten wurde in Farbe gescannt. Hierbei nimmt der Scanner zusätzlich Fotos auf und überträgt diese Farbinformationen auf jeden einzelnen Scanpunkt. Voraussetzung dafür ist eine gute Ausleuchtung des Messbereiches durch externe Lichtquellen.

Zusammenfassend sind beide Laserscanner den Anforderungen gewachsen und lieferten Punktwolken mit einem erwartbaren, durchschnittlichen Punktrauschen von wenigen Millimetern. Auch die zusätzlichen Farbbilder wurden mithilfe intensiver externer Beleuchtung in guter Schärfe und Kontrast erstellt.

Als Ansatz für die Volumenberechnung wurde die Erzeugung eines “wasserdichten” Meshes aus den kombinierten Punktwolken vor und nach der Sprengung gewählt. Dazu wird zunächst ein Mesh der Ortsbrust nach der Sprengung inkl. ihres Vorraums erstellt und im zweiten Schritt mit dem Mesh der Ortsbrust vor der Sprengung “verschlossen” (Abb. 6). Von dem so entstandenen Volumenkörper lässt sich die Menge des herausgesprengten Materials ableiten.

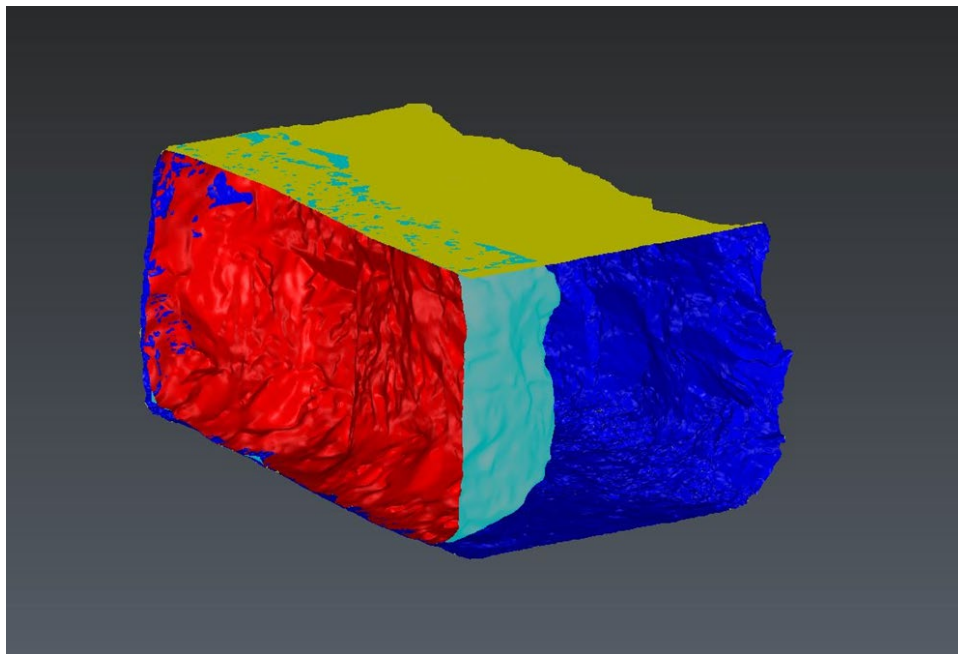


Abb. 6: Seitenansicht der verbundenen Meshes (rot: Ortsbrust vS / blau: Ortsbrust nS)

Das Ergebnis dieser Methode stellt in jedem Fall einen Näherungswert dar, denn im Zuge der Erstellung bedarf es zum einen einer Generalisierung der Mesh-Oberflächen und zum anderen der Interpretation des Bearbeiters.

Die sehr unregelmäßige und verwinkelte Felsoberfläche führt bei Verwendung der kompletten Punktwolke zu Löchern und Fehlern im Mesh, deren Behebung mit größerem manuellem Aufwand verbunden wäre. Ein vorab angewendetes Ausdünnen bzw. Filtern der Punktwolke ermöglicht hier eine schnellere und vor allem lückenlose, automatische Berechnung des Meshes. Hierfür hat sich ein Punktabstand von 20 cm als zielführend erwiesen.

Ein gewisser Interpretationsspielraum ergibt sich außerdem bei der Segmentierung der Punktwolke, also der Auswahl der jeweiligen Bereiche, die für das Mesh berücksichtigt werden. Kritisch ist hier beispielsweise der Übergang zwischen der Ortsbrust vor der Sprengung und der Punktwolke in diesem Bereich nach der Sprengung. Des Weiteren können größere Lücken in der Ortsbrust vorhanden sein, wenn diese sich in einen Kreuzungsbereich mit weiteren Nebenorten befindet. Derartige Löcher im Mesh müssen dann durch Interpolation manuell verschlossen werden.

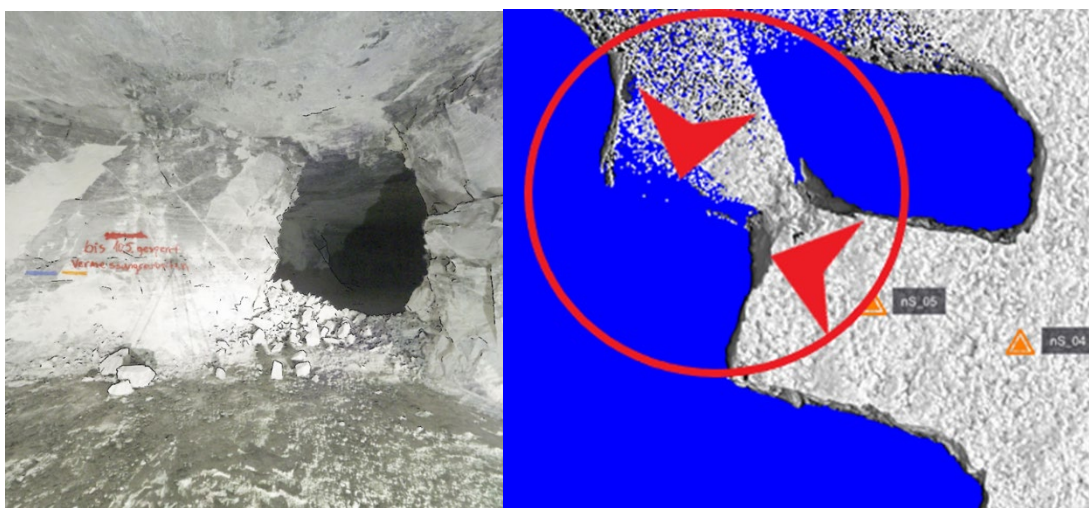


Abb. 7: Blick auf die Ortsbrust nS mit Durchbruch zum Nebenort (links) / Durchbruchssituation in der Draufsicht mit Vortriebsrichtungen (rechts)

## 2.2 Optimierung der Bohrtiefen

Aufgrund der geologischen Gegebenheiten und des Bohrschemas entstehen nach jeder Sprengung Unebenheiten an der neuen Ortsbrust. Erneuter Vortrieb mit gleichbleibenden Bohrtiefen führt zur Zunahme dieser Unebenheiten und einem ineffizienten Gesteinsabbau, weil die Ortsbrust mechanisch nachgearbeitet werden muss. Durch die Optimierung der Bohrtiefen wird beabsichtigt nach einer Sprengung eine neue, ebene Ortsbrust zu erreichen. Durch den Abschluss aller Bohrungen auf einer zur Vortriebsrichtung senkrecht stehenden Ebene soll dies gewährleistet werden. Das Ziel ist es die individuelle Tiefe aller Bohrlöcher auf Basis der Laserscanning-Daten zu berechnen. Dies soll auch eine Optimierung der Bohrarbeit und der eingesetzten Sprengstoffmenge ermöglichen.

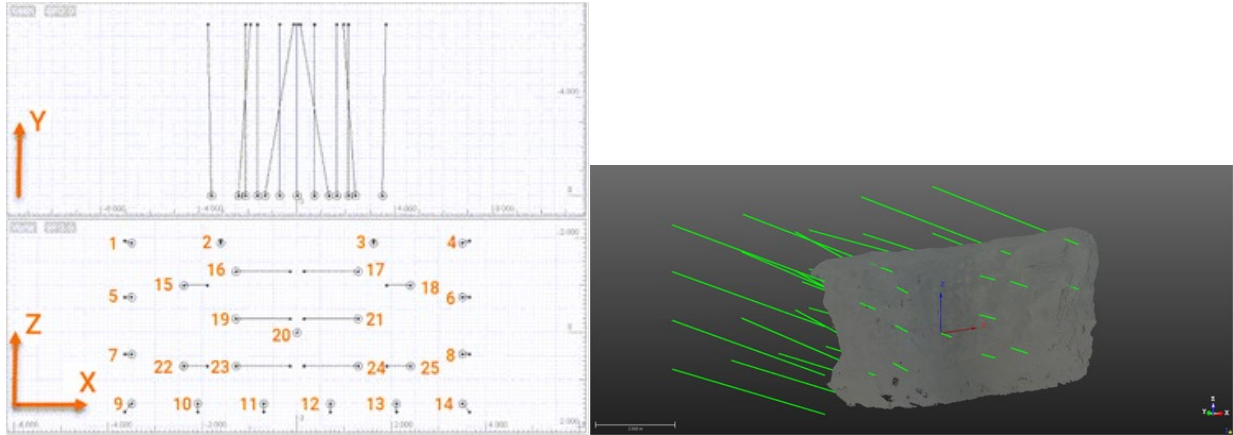


Abb. 8: Bohrlochplan und Visualisierung der Bohrlochtiefen in Abhängigkeit der aktuellen Ortsbrust

Die Bohrungen orientieren sich an der “Zeigerschicht”, die händisch auf der Ortsbrust durch Sprühfarbe markiert wird. Zur Ermittlung der optimalen Bohrlochtiefen wird die Ortsbrust mit dem TLS gescannt. Der relevante Teil der entstehenden Punktwolke wird ausgeschnitten und so orientiert, dass die Vortriebsrichtung in y-Achsenrichtung zeigt. Der Sprühmarkierung dient als Ursprung des Koordinatensystems. Die x,z-Ebene steht senkrecht zur Vortriebsrichtung, also zur y-Achse. Das 3D-Bohr raster (Abb. 8 links) wird im lokalen Koordinatensystem eingepasst. Entlang der Vortriebsrichtung werden die Bohrlöcher unter Berücksichtigung des Bohrwinkels bis zu einer vorgegebenen einheitlichen Tiefe visualisiert und mit der gemessenen Ortsbrust verschnitten. Dadurch ergeben sich die in Abb. 8 (rechts) gezeigten aus der Ortsbrust herausstehenden Linienstücke, um deren Länge die jeweilige Bohrung gekürzt wird, damit die Bohrlochenden nach Abschluss aller Bohrungen in einer Ebene liegen. Die Untersuchungen zur Optimierung des Bohrrasters erfolgten in Trimble Realworks sowie in CloudCompare.

Das Testszenario hat gezeigt, dass die Optimierung der Bohrlochtiefen aus 3D-Punktwolken möglich ist. Für eine effiziente Umsetzung ist eine automatische Maschinensteuerung notwendig.

### 2.3 Dokumentation des Abbaus

Im Rahmen der Verordnung über markscheiderische Arbeiten und Beobachtungen der Oberfläche (Markscheider-Bergverordnung - MarksBergV), ist der Betreiber des Gewinnungsbetriebes dazu verpflichtet den Abbau zu dokumentieren. Hierfür wurde ein Ansatz zur Dokumentation des Abbaus mit terrestrischem Laserscanning ausgearbeitet und das o.g. Festpunktfeld (Kap. 2.1.2) verwendet. Mithilfe der target-basierten Registrierung ist es möglich die Scans (automatisch) im übergeordneten Koordinatensystem zu verorten, zu georeferenzieren. Die Abbaugrenzen sind in den Laserscans ersichtlich und können lagegenau in das Risswerk / den Grubenplan übernommen werden.

Mit heutigen Softwarepaketen ist es des Weiteren möglich eine Punktwolke der gesamten Grube vorzuhalten, d. h. in einem dreidimensionalen GIS (Geoinformationssystem) die Geometrie zu dokumentieren und laufend fortzuschreiben. Aufgrund der Größe des Gebietes ist allerdings eine verzerrungsfreie Abbildung in der Ebene mit Maßstab 1 nicht möglich. Andererseits ist es in den gängigen Softwareprodukten zur Punktwolkenbearbeitung und -speicherung nicht möglich geodätische Korrekturen für Projektion und Höhe anzubringen. Mit der Verwendung der Passpunkte im

übergeordneten Koordinatensystem– in diesem Fall einer Gauß-Krüger-Projektion– werden die einzelnen Punktwolken mittels einer Affinen Transformation georeferenziert. Fehler aus den nicht reduzierten Strecken kommen somit nur innerhalb einer einzelnen Punktwolke vor und sind aufgrund der geringen räumlichen Ausdehnung vernachlässigbar. In der Gesamtsicht, der Positionierung einzelner eigenständig georeferenzierter Punktwolken im übergeordneten System, ist die Abbildung aber korrekt.

In dem Gruben-GIS können Veränderungen über die Zeit dokumentiert aber auch (flächenhafte) Deformationen ermittelt werden, wenn Bereiche zeitlich versetzt mehrfach erfasst werden.

### 3 Fazit & Ausblick

Im Laufe des Projektes konnten die Berechnung des Abbauvolumens, die Sicherstellung einheitlicher Bohrtiefen sowie die Dokumentation des Abbaus erfolgreich exemplarisch realisiert werden. Die Optimierungsansätze wurden mit verschiedenen Softwarepaketen für die 3D-Punktwolkenauswertung erprobt.

Bei den eingesetzten Programmen wurden Arbeitsabläufe unter Berücksichtigung der Integrierbarkeit in vorhandene Prozessketten durchlaufen und dahingehend kritisch beurteilt. Dabei stellte sich eine organisationstechnische Zweistufenlösung als zielführend heraus. Die 3D-Datenerfassung, -auswertung und -weiterverarbeitung soll durch firmeninternes Personal seitens des Bergbauunternehmens bewerkstelligt werden. Die Schulung des Personals des Bergbauunternehmens in Messung und Auswertung von Laserscanningdaten für wiederkehrende Abläufe, sowie Grundlagenstellung und Integration in die Prozessketten ist extern beabsichtigt.

Weitergehende Ziele sind die Weitergabe des abgeleiteten Bohrrasters an die Maschinensteuerung des Bohrwagens zur automatisierten Einstellung der optimalen Position der Bohrlöcher und der Bohrtiefen. Die Positionierung und Ausrichtung des Bohrwagens, kann dabei mithilfe eines im Festpunktfeld stationierten Tachymeters maschinengestützt/automatisiert erfolgen. Die Erkennung der “Zeigerschicht” (Charakteristische Gesteinsschicht in der Lagerstätte) im digitalen Datensatz (ggfls. IR-Bild des Laserscanners) und die damit verbundene automatische Höheneinstellung des Bohrrasters, die Optimierung der Bohrtiefen mit dem Ziel der Reduktion der Bohrungen und des Sprengstoffverbrauchs sind für Folgearbeiten vorgesehen.

### 4 Literaturverzeichnis

- ARCH COAL, 2023: ABC 123: Testing Site / Course zitiert nach Arc Resources, Inc., <https://www.e-education.psu.edu/geog000/node/900>, letzter Zugriff 19.02.2023.
- DEGEN, I., WEBER, L. & BRUNN, A., 2014: Erstellung eines 3D-Höhlenmodells einer Tropfsteinhöhle. Publikationen der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation e.V., **23**, Seyfert, E., Gülch, E., Heipke, C., Schiewe, J. & Sester, M. (Hrsg.).
- HÖBELBART, D.; RICHTER, O., MARTIN, C., MARTIENßEN, TH. & WANDINGER, M., 2018: Erste Erfahrungen mit dem handgeführten Laserscanner ZEB-REVO bei Vermessungsarbeiten im Altbergbau am Südtiroler Schneeberg. 18. Altbergbau-Kolloquium, Wieliczka.

- JAWECKI, B., DABEK, P.B., PAWESKA, K. & WIE, X., 2018: Estimating Water Retention in Post-mining Excavations using LIDAR ALS Data for Strzelin Quarry in Lower Silesia. *Mine Water and the Environment*, **37**, 744-753.
- KAJZAR, V., KUKUTSCH, R. & HEROLDOVA, N., 2015: Verifying the Possibilities of using a 3D Laser scanner in the Mining Underground. *Acta Geodyn. Geomater.*, 1/12 (177), 51-58.
- KUCHENBECKER, R., 2010: Sub-Terrestrisches Laserscanning – Anwendungsbeispiele aus dem Bergbau unter Tage. DVW-Schriftenreihe, **64**, TLS 2010.
- LAUX, D., 2017: Terrestrisches Laserscanning zur Quantifizierung von Trennflächenparametern für statistisch signifikante Discrete-Fracture-Network-Modelle. Dissertation an der TU Darmstadt.
- MATWIJ, W., GRUSZCZYNSKI, W., PUNIACH, E. & CWIAKALA, P., 2021: Determination of Underground-mining Displacement Fields using Multitemporal TLS Point Cloud Registration. *Measurement*, **180**.
- OPARIN, V.N., SEREDOVICH, V.A., YUSHKIN, V.F., IVANOV, A.V. & PROKOP'eva, S.A., 2007: Application of Laser Scanning for Developing a 3D Digital Model of an Open-pit Side Surface. *Journal of Mining Science*, **43**(5).
- PEJIC, M., 2013: Design and Optimisation of Laser Scanning for Tunnels Geometry Inspection. *Tunnelling and Underground Space Technology*, **37**, 199-206.
- STUDNICKA, N. & GROISS, B., 2019: Untertage-Aufnahme und anschließende Demokratisierung von terrestrischen Laserscandaten. 20. Geokinematischer Tag, Freiberg, 172-181.
- TSCHARF, A., 2018: Stand der Technik in der Tagbauvermessung. *Berg- und Hüttenmännische Monatshefte*, **163**, 57-63.
- WUJANZ, D., BARAZETTI, L., PREVITALI, M. & SCAIONI, M., 2019: A comparative study among three registration algorithms: Performance, quality assurance and accuracy. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, **42**(2/W9), 779-786.

# Detektion großräumiger Niedrigwassersituationen durch satellitenbasierte Wasser-Land-Grenzen-Bestimmung („Sat-Land-Fluss“)

BASTIAN GESSLER<sup>1</sup>, SILKE MECHERNICH<sup>1</sup> & BJÖRN BASCHEK<sup>1</sup>

*Zusammenfassung: Ausdehnung und Lage von wasserbedeckten Gebieten sind für vielfältige Anwendungen von Interesse wie z.B. für das Monitoring von hydrologischen Extremereignissen. Um die auftretenden Prozesse besser verstehen und Schäden möglichst gering halten zu können, sind quantitative Informationen über Lage und Verlagerung der Grenze zwischen Wasser und Land von großer Bedeutung (Wasser-Land-Grenze, WLG). Der Radarsatellit Sentinel-1 (S-1) bietet hier großes Potential, deshalb verwendet das Projekt „Sat-Land-Fluss“ diese Daten als Grundlage. Bei der Auswertung der S-1 Daten sind technische Herausforderungen zu lösen und eine Kombination mit anderen Daten zielführend, um Datenlücken und Unsicherheiten zu minimieren. Im Projekt werden Radar-Rückstreuungsinformationen aufbereitet und mit verschiedenen Geobasisdaten kombiniert. Schwerpunktmäßig untersucht das Projekt, wie weit die Integration dieser Zusatzdaten die Qualität der WLG-Produkte steigern kann. Dieser Beitrag fokussiert sich auf die Niedrigwassersituation 2018 am Rhein.*

## 1 Einleitung und Motivation

Hydrologische Extremereignisse sowie kontinuierliche Prozesse verändern laufend die Grenze zwischen Wasser und Gelände bzw. den trockengefallenen Flächen der Wasserwechselzone. Die Ausprägung und Verlagerung der Wasser-Land-Grenze (WLG) sind wesentliche Größen für das Prozessverständnis im Bereich der Binnen- und Küstengewässer. Darum ist deren Erfassung bzw. Monitoring eine Basisvoraussetzung zur Bearbeitung vielfältiger Fragestellungen des Gewässermanagements.

Dazu bietet sich aus fachlichen und wirtschaftlichen Überlegungen der Einsatz satellitengestützter Fernerkundungsverfahren, insbesondere die Nutzung der kostenfreien Copernicus-Datenprodukte des Satelliten Sentinel-1 (S-1, Radar), an. Das Projekt „Sat-Land-Fluss“ zielt darauf ab, die Verlässlichkeit der aus S-1 Daten abgeleiteten WLG zu analysieren und dabei das Potential, sowie die Limitierungen der Produkte zu untersuchen.

Dieser Beitrag zeigt auf, wie weit die Integration von Zusatzdaten die S-1 Datenauswertungen in Niedrigwassersituationen verbessern. Solche Ableitungen können z.B. für das Monitoring von Extremereignissen, ein Veränderungsmonitoring zur Unterstützung des Sedimentmanagement und zur Überwachung der Schiffbarkeit dienen.

---

<sup>1</sup> Bundesanstalt für Gewässerkunde (BfG), Am Mainzer Tor 1, D-56068 Koblenz, E-Mail [gessler, mechernich, baschek]@bafg.de

## 2 Forschungsstand

Die Nutzung von Radardaten (SAR) bietet sich an, da die Daten – im Gegensatz zu optischen Satellitendaten - unabhängig von Tageslicht und Wetter auswertbar sind. Entsprechend haben die Radardaten ein höheres Potential, um kurzfristige Ereignisse zeitgerecht zu detektieren (OLTHOF & TOLSZCZUK-LECLERC 2018).

Aufgrund der schrägen Aufnahmegeometrie kommt es bei SAR Bildern zu geometrisch bedingten Abbildungseffekten, wie *Foreshortening*, *Layover* und Abschattung, die durch Relief, Bauwerke und hohe Vegetation verursacht werden. Während *Foreshortening* und *Layover* durch geometrische Korrekturen weitestgehend behoben werden können, treten im Bereich von Abschattungen (z.B. durch Steilkanten) Datenlücken auf. Um die Datenlücken zu schließen, lässt die Nutzung von Zusatzdaten (z.B. DGM-Integration) erhebliche Produkt-Verbesserung erhoffen (HUANG et al. 2018; LIAO & WEN 2020; OLTHOF & TOLSZCZUK-LECLERC 2018).

Meist wird zur WLG-Ableitung die Intensität der Rückstreuung genutzt, da glatte Wasseroberflächen spiegelnde Eigenschaften und somit geringere Rückstreuintensitäten aufweisen (MARTINIS et al. 2015). Die Rückstreueigenschaft von Gewässeroberflächen erhöht sich bei rauen Wasseroberflächen, die bei Strömung, Schiffswellen oder zunehmender Windstärke entstehen. Glatte Landoberflächen (z.B. Flugplätze) oder durchfeuchtete Gebiete weisen hingegen erheblich reduzierte Rückstreuungen auf, so dass eine dem Wasser vergleichbare spiegelnde Eigenschaft zu verzeichnen ist (MARTINIS et al. 2018; KUMAR 2021). Um Interpretationsfehler zu vermeiden werden die SAR-Daten häufig mit Informationen anderer Sensoren, wie optischen Daten oder digitalen Gelände- und Oberflächenmodellen kombiniert oder eine multitemporale Auswertung von Daten durchgeführt (OLTHOF & TOLSZCZUK-LECLERC 2018; LIAO & WEN 2020; HUANG et al. 2018; MARTINIS et al. 2018).

Verschiedene Projekte und laufende Dienste zeigen die generelle Machbarkeit der WLG-Ableitung aus satellitengestützten Radardaten, z. B. existieren mehrere Dienste des Copernicus Emergency Management Service (CEMS; Rapid Mapping, Risk and Recovery Mapping, Global Flood Monitoring), die Satellitendaten zur Erstellung von Hochwasserprodukten (inkl. WLG) verwenden. Die beiden Kartierungsarten des CEMS-Dienstes stellen in/nach einer Krise Karten bereit, die auf Auswertungen von Satellitenaufnahmen basieren (BBK 2023). Die Dienste müssen über eine nationale Kontaktstelle aktiviert werden. Beim *Global Flood Monitoring* (GFM) werden dagegen automatisiert alle S-1 Szenen ausgewertet und jeweils eine abgeleitete Überflutungsfläche, sowie weitere Informationen (u.a. Ausschlusswerte, Unsicherheiten) in einer Webanwendung visualisiert (EOC 2023). Die ausgewerteten Produkte zeigen in der aktuellen beta-Version Plausibilitätsfehler bei der Wasserklassifizierung auf trockengefallenen Flächen auf. Durch die oftmals fehlende Validierung ist die Belastbarkeit der Produkte weitgehend unbekannt und somit die Nutzung für kritische Fragestellungen nur eingeschränkt möglich.

Im Gegensatz zu den operationellen Produkten, nutzt dieser Beitrag hochaufgelöste Geländeinformationen zur Datenvorprozessierung (S-1) und verschiedene GeoBasisdaten zur Erstellung von Ausschlussmasken in Zuge der WLG-Bestimmung. Die abgeleiteten Wasserflächen werden mit hochaufgelösten, zeitlich in geringem Abstand versetzten Validierungsdaten abgeglichen, wodurch eine präzise Validierung vorliegt.



## 3 Datengrundlagen

### 3.1 Sentinel-1

Bei den verwendeten S-1 Daten handelt es sich um Level-1 SLC (Single Look Complex) Daten, die im *Interferometric Wide Swath* Modus aufgezeichnet werden und sowohl VV als auch VH-polarisiert sind. Die räumliche Auflösung beträgt bei S-1 beträgt circa 5 x 20 m. In Deutschland deckt der verbleibende S-1A ein Gebiet ca. alle 6 Tage ab (ESA 2022).

Bei der Auswertung von Radardaten sind die in Kap. 2 genannten technische Herausforderungen zu lösen. Zudem stößt die Erfassung von bestimmten Objekten aufgrund der relativ niedrigen räumlichen Auflösung teilweise an ihre Grenzen, z. B. bei der Erfassung „kleiner“ Strukturen wie Bühnen. Diese können je nach Blickrichtung des Satelliten, Einfallswinkel der Radarstrahlen bzw. Größe der Objekte nicht immer erfasst werden (JENEROWICZ et al. 2019), da die Radar-Rückstreuung vom Einfallswinkel abhängt und somit u. a. je nach Satelliten-Orbit (aufsteigend/absteigend) und Entfernung eines Objekts zum Sensor variiert (KAPLAN et al. 2021).

### 3.2 Zusatzdaten

Die verwendeten Zusatzdaten zur Verbesserung der WLГ-Ableitung aus S-1 Daten werden im Folgenden kurz vorgestellt.

#### a) Landbedeckung/Landcover – ESA WorldCover 2020

Im Projekt werden Landbedeckungsinformationen des frei und global verfügbaren Datensatzes *WorldCover* der Europäischen Weltraumorganisation (ESA) verwendet (Version 2020). In den WorldCover Daten wurden optische (S-2), sowie Radardaten (S-1) kombiniert (ESA 2023a).

#### b) Höhendaten (DGM5) und HAND-Index

Aus dem Digitalen Geländemodell Gitterweite 5 m (DGM5, 2016) des Bundesamt für Kartographie und Geodäsie (BKG) wurde im Projektverlauf der HAND-Index (Height Above Nearest Drainage, NOBRE et al. 2011) berechnet. Während der Prozessierung wird die Geländetopographie entlang eines bestimmten Entwässerungssystems in Beziehung zur relativen Höhe gesetzt, so dass der Index das Entwässerungspotential des jeweiligen Pixels widerspiegelt.

#### c) Fahrrinneninformationen

Im Bereich der Fahrrinne der Bundeswasserstraßen liegt stets Wasser vor, auch im extremen Niedrigwasserfall. Die Lage der Fahrrinne wurde aus den Daten des Wasserstraßen- und Schifffahrtsamtes (WSA) Rhein entnommen (Stand 2021).

### 3.3 Validierungsdaten

Um beurteilen zu können, wie gut der Algorithmus die WLГ abbildet, sind Validierungsdaten erforderlich. Für die in diesem Beitrag ausgewertete Niedrigwassersituation am Mittelrhein 2018 (Abb. 1) wurden Informationen aus einem Bildflug verwendet, der im Auftrag des Wasserstraßen- und Schifffahrtsamtes (WSA) Oberrhein durchgeführt wurde.

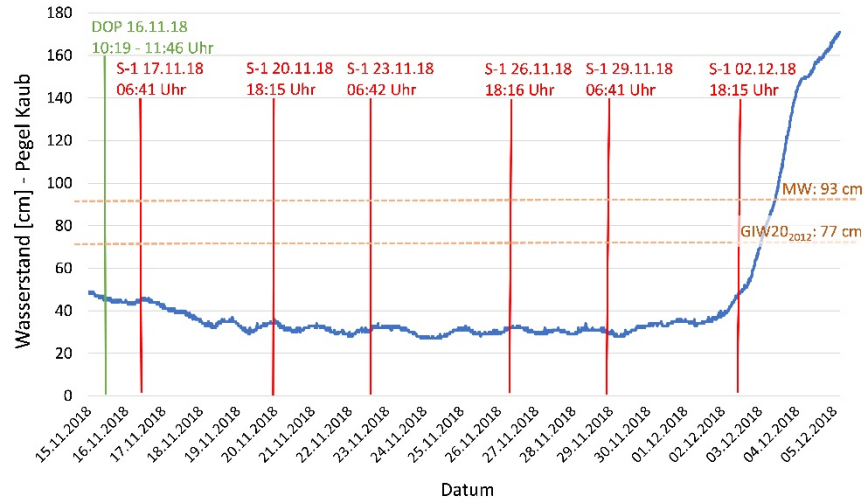


Abb. 1: Wasserstände um die Niedrigwassersituation 2018 am Pegel Kaub (Mittelrhein), mit den Zeitpunkten der verfügbaren Sentinel-1 Daten (rot) sowie der Befliegung (grün). Der Mittelwasserstand (MW) und gleichwertiger Wasserstand (GIW202012, Wasserstand der statistisch an 20 Tagen im Jahr unterschritten wird) sind als horizontale orange Linien abgebildet

Der Bildflug fand im Projektgebiet am 16.11.2018 statt, und somit kurz vor dem Zeitpunkt des niedrigsten Wasserstandes 2018. Aus den Bildflugdaten generierte die BfG DOPs mit 10 cm Bodenauflösung, woraus in einem GIS die WLG visuell bzw. manuell abgeleitet wurde.

## 4 Methodik

Das methodische Vorgehen bei der Datenprozessierung ist in Abb. 2 dargestellt. In die Aufbereitung der S-1 Daten (Vorprozessierung) fließen Höheninformationen des DGM5 ein, die bei der Geländekorrektur für die Homogenisierung der Daten sorgen. Bei der Reduzierung des sogenannten *Speckle*-Effekts werden Filter verwendet, die Objektstrukturen nach Möglichkeit optimal zur späteren Extraktion der Wasserflächen konservieren. Der *Speckle*-Effekt muss deshalb je nach Anwendung möglichst stark gemindert werden, da dieser Störeffekt den Informationsgehalt und die Bildinterpretation negativ beeinflusst (KUMAR 2021). Die Vorprozessierung ist nach weitestgehender Bereinigung der S-1 Daten abgeschlossen und es folgt die Grenzwertbildung nach KITTNER & ILLINGWORT (1985).

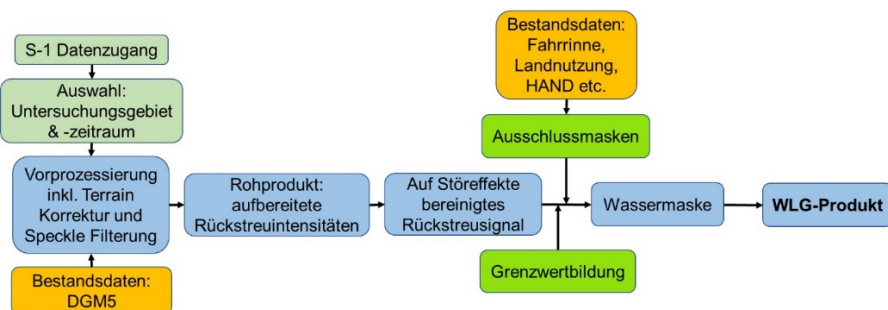


Abb. 2: Schematische Darstellung der Methodik der WLG-Ableitung

Anschließend werden mehrere Ausschlussmasken erstellt, die in Bereichen von sicherem Wasser (Fahrrinne) oder sicherem Land (bestimmte Landnutzungsklassen und Entwässerungspotentiale) angewandt werden. Aus der Kombination der ausgewerteten S-1 Daten und den Masken der Zusatzdaten ergibt sich eine Wassermaske (Abb. 2). Aus dieser können im weiteren Projektverlauf die WLГ sowie weitere WLГ-Produkte errechnet werden.

## 5 Anwendungsbeispiel Niedrigwasser Mittelrhein

Das Projektgebiet für das Niedrigwasser 2018 wurde im Bereich des Mittelrheins gewählt (Abb. 3). Hier befinden sich mehrere Kiesbänke, von denen einige immer und andere nur bei niedrigen Wasserständen trockenfallen. Die starke Topographie (Abb. 3b) erschwert die vollständige Datenerfassung aufgrund von Abschattungen, so dass hier die aufwändige Geländekorrektur während der S-1 Datenaufbereitung einen Mehrerfolg erzielt.

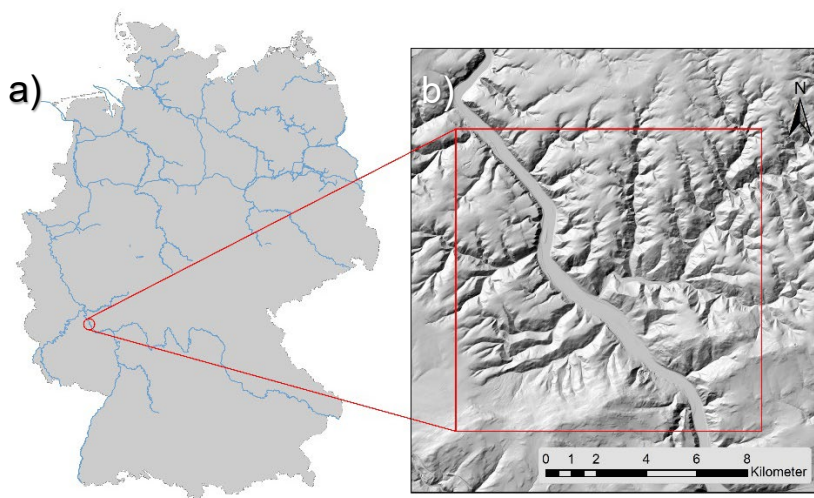


Abb. 3: (a) Deutschland mit dem Verlauf der Bundeswasserstraßen (BWA, © WSV) und dem Punkt des Untersuchungsgebietes, (b) Schummerung im Bereich des Mittelrheins (© basemap.de / BKG Januar 2023)

Am Rhein begann die Niedrigwassersituation 2018 im August und hielt auch noch den kompletten November bei sehr ähnlichen Wasserständen an. Im November fand am Mittelrhein ein Bildflug und sechs Sentinel-1 (S-1) Aufnahmen statt (Abb. 1). Der Bildflug und die daraus abgeleitete WLГ diente zur Validierung der S-1 Auswertungen. Die S-1 Szene vom 17.11.2018 wird hier als Beispiel gezeigt. Nach eingehender Prüfung beider Polarisierungen (VH, VV) der S-1 Szene des 17.11.2018, wurde die VV-polarisierten Daten als weniger fehleranfällig identifiziert. Deshalb beziehen sich alle weiteren Abbildungen auf diese Polarisierung.



Abb. 4: Niedrigwassersituation am Mittelrhein am 17.11.2018, (a) Sentinel-1 Rückstreuintensität (je dunkler, desto weniger Rückstreuung des Radarsignals), (b) Wassermaskenableitung aus S-1 Daten ohne Zusatzdaten, (c) Wassermaskenableitung aus S-1 Daten und HAND-Algorithmus, (d) Wassermaskenableitung aus S-1- und Fahrrinnendaten, (e) Wassermaskenableitung aus S-1- und Landbedeckungsdaten, (f) Wassermaskenableitung aus S-1- sowie allen Zusatzdaten. Die Buchstaben L (Landbedeckungsdaten), H (HAND-Algorithmus) und F (Fahrrinnendaten) symbolisieren die exemplarisch aufgezeigten Verbesserungen durch die entsprechenden Zusatzdaten

In Abb. 4a ist die Rückstreuintensität der aufbereiteten S-1 Szene dargestellt. Die dunkelsten Pixelwerte werden dabei in der Mitte des Bildes erreicht. Hierbei handelt es sich um den Rhein, bei dem aufgrund seiner relativ glatten Wasseroberfläche ein Großteil der vom Satelliten ausgesendeten Strahlung nicht mehr zum Sensor zurückgestreut wird. In den folgenden Abbildungen 4b-f ist die WLK aus dem Bildflug als rote Linie dargestellt.

### **5.1 WLK durch ausschließliche Nutzung der S-1 Daten**

Die nach der Vorprozessierung abgeleiteten Wasserflächen aus den S-1 Daten zeigen den Rhein als zusammenhängende blaue Struktur (Abb. 4b). Zusätzlich sind etliche Wasserflächen bzw. größere Anhäufungen von Pixeln abseits des Flusses auszumachen, bei denen es sich um Fehlklassifizierungen handelt. Diese werden durch Grauwerte, die den Pixelwerten der tatsächlichen Wasserflächen im Bereich des Gewässers zu ähnlich sind, hervorgerufen.

### **5.2 Ableitung inklusive der Zusatzdaten mittels HAND-Index**

Durch die Berechnung des HAND-Index aus den hochaufgelösten Höheninformationen wurde der Ausschluss von Flächen ermöglicht, die aufgrund ihrer topographischen Lage als nicht wasserbedeckt eingeordnet werden können (Abb. 4c). In Abb. 4c sind größere Bereiche von Wasser-Pixeln außerhalb des Rheins, im Vergleich zu den S-1 Wasserflächen ohne Zusatzdaten (Abb. 4b), verschwunden. Fehlklassifizierungen in Abb. 4c bestehen größtenteils noch in Tallagen, die von der HAND-Ausschlussmaske nicht abgedeckt werden, da der Index diese als potentiell wasserbedeckt einstuft.

### **5.3 Ableitung inklusive der Zusatzdaten Fahrrinnenlage**

Verbesserungen durch Fahrrinneninformationen stellen sich im abgebildeten Untersuchungsgebiet als relativ gering dar, da während der Aufzeichnung der S-1 Szene scheinbar kaum Überlagerungseffekte von Schiffen stattgefunden haben, nur kleine Metallstrukturen (Bauwerke) im Bereich der Fahrrinne vorliegen und die Wasseroberfläche glatt war. Ein Verbesserungseffekt durch Fahrrinnenlage-Information ist somit stark auf die Anwesenheit dieser Objekte und den Windbedingungen bzw. der Windexponiertheit beschränkt. In weiter nördlich gelegenen Bereichen des Rheins wurden deutliche Verbesserungen der Wassermaske durch die Integration der Fahrrinnenlage nachgewiesen.

### **5.4 Ableitung inklusive der Zusatzdaten Landbedeckung**

Durch die Integration der Landbedeckungsdaten ist in Abb. 4e, ähnlich wie in Abb. 4c, deutlich zu beobachten, dass große, durch uneindeutige Rückstreuungseffekte fälschlicherweise als Wasser klassifizierte Gebiete, nun richtig zugeordnet werden. Die Hintergrundinformation des DOP20 lässt deutlich werden, dass an den Stellen der entsprechenden Pixel homogene Landflächen anzutreffen sind (Abb. 4e). Hierbei handelt es sich beispielsweise um landwirtschaftliche Anbauflächen oder Wiesen. Aufgrund der vorherrschenden niedrigen Vegetation, des gepflügten Bodens etc. ist in diesen Gebieten die Rückstreuung relativ gering und es kommt durch das Schwellwertverfahren zur fehlerhaften Klassifizierung. Aber auch Abschattungsbereiche im Umfeld größerer Erhebungen lösen vermutlich die niedrigen Grauwerte aus, denn hier interpoliert die Geländekor-

rektur während der Vorprozessierung der S-1 Daten über den Radarschatten hinweg. Die abgeschatteten Bereiche werden wahrscheinlich durch angrenzende Pixel, die relativ geringe Rückstreuwerte aufweisen, gemittelt. Letztendlich kommt es zu den vermeintlich als Wasser klassifizierten Bildelementen. Verbleibende Fehlklassifizierungen finden sich noch in Randbereichen bestimmter Landnutzungsklassen, die vermutlich nicht oder nur teilweise in den Bereich der Ausschlussmaske fallen.

## 5.5 Validierung und Interpretation

Die Validierung der abgeleiteten Wasserflächen erfolgt sowohl visuell (Abb. 4, 5) als auch quantitativ (Tab. 1-3).

Die räumlichen Diskrepanzen zwischen den Wasserflächen aus den Befliegungsdaten und den abgeleiteten Wasserflächen aus den S-1 Daten inklusive aller Zusatzdaten sind in Abb. 5 dargestellt. Es wird deutlich, dass die meisten Fehlklassifizierungen in direkter Umgebung zum Rhein verbleiben. Speziell befinden sich die meisten falsch klassifizierten Wasserbereiche vermehrt am linksrheinischen Ufer, konzentriert im Bereich von Buhnen oder in deren näheren Umgebung. Wahrscheinlich sind Abschattungseffekte aufgrund der Geländetopographie und der Aufnahmegeometrie (absteigender Orbit, *descending*) dafür verantwortlich. Außerdem sind die meisten Buhnen zu klein um vom S-1 aufgelöst werden zu können. Des Weiteren konzentrieren sich falsch klassifizierte Wasserpixel auf trockengefallene Bereiche der Kiesbänke. Hier könnte eine Durchfeuchtung die niedrige Radar-Rückstreuung hervorgerufen haben. Vereinzelt finden sich Pixelgruppen an den Hängen der Seitentäler des Rheins, bei denen ebenfalls Abschattungseffekte als Ursache angenommen werden. Insgesamt liegt im Abstand von ca. 50-100 m um die Uferlinie eine unsichere Klassifizierung vor (Abb. 5). Abseits der direkten Umgebung des Rheins weisen große Gebiete eine korrekte Klassifizierung vor.

Für die quantitative pixelweise Evaluierung wird beispielhaft die Konfusionsmatrix der Klassifizierungsergebnisse unter Verwendung aller Zusatzdaten sowohl nach absoluten (Tab. 1) als auch nach prozentualen Anteilen (Tab. 2) dargestellt. Eine Gesamtübersicht der Qualität der jeweiligen Klassifizierungen wird in Tab. 3 betrachtet.

Die Zeilen der Konfusionsmatrizen (Tab. 1, 2) geben die abgeleiteten Klassen aus den S-1 Daten wieder und in den Spalten werden die aus dem Bildflug berechneten Klassifizierungsergebnisse erfasst. Die Zahlen in der obersten, linken Zelle von Tab. 1 geben die Pixelanzahl der Land-Klasse wieder, die aus den S-1 Daten als Land-Pixel bestimmt wurden und auch in den Validierungsdaten (DOP) dieser Klasse zugeordnet wurden. Diagonal nach unten versetzt ist dies für die Klasse *Wasser* erfolgt. Die nicht fettgedruckten Werte geben die aus den S-1 Daten fälschlicherweise als Wasser bzw. Land eingeordneten Pixel wieder. Der prozentuale Anteil an der Gesamtpixelanzahl wird in Tab. 2 dargestellt.

Wie bereits in der visuellen Analyse festgestellt, wird auch in der tabellarischen Analyse der verhältnismäßig große Anteil korrekt klassifizierter Land-Pixel (Tab. 1, 2: 1.491.101 Pixel, 96,42 %) deutlich.

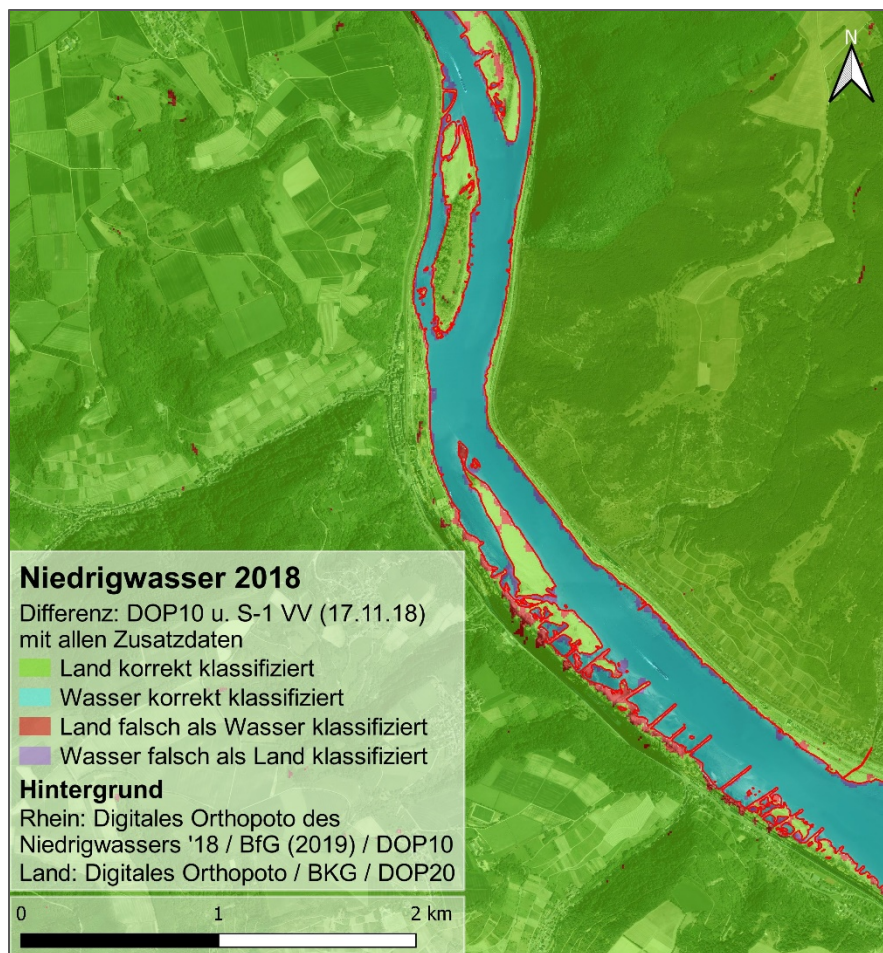


Abb. 5: Über Hintergrundbilddaten (DOP10 und DOP20) werden die Unterschiede zwischen den Wasserflächen aus den Befliegungswasserflächen und den abgeleiteten Wasserflächen aus den S-1 Daten inklusive aller Zusatzdaten dargestellt

Tab. 3 zeigt die Auswertungen der verschiedenen Algorithmen, um das Verbesserungspotential der Zusatzdaten beurteilen zu können. Werden die Ergebnisse der S-1 Ableitung auf Pixelebene analysiert bestätigt Tab. 3, dass bei den erfolgreich klassifizierten Land-Pixel bis ~96 % erreicht werden. In direkten Zusammenhang dazu steht der Anteil falsch als Wasser klassifizierte Pixel, da diese Klasse sich durch die Zusatzdaten deutlich verringert und die meisten zuvor falsch als Wasser klassifizierten Pixel zu Land-Pixel korrigiert werden. Gemessen an der Gesamtpixelanzahl liegt dieser ohne Zusatzdaten ~9 % höher als mit allen Zusatzdaten (0,42 %, Tab. 3).

Die größte Verbesserung wird mittels zusätzlicher Informationen durch den Einsatz von Landbedeckungsdaten und des HAND-Index erreicht (Tab. 3). Dies gilt insbesondere für die fehlerhaft als Wasser klassifizierten Flächen. Die Werte korrekt eingeordneter Wasser-Pixel stagnieren durchweg bei ~2,9 %, da sich nur wenige Pixel durch die Verwendung der Zusatzdaten verbessern. Dies zeigt, dass Verbesserungen im Bereich des Rheins nicht signifikant durch die Zusatzdaten erreicht werden. Wenig Veränderung besteht ebenfalls bei der fälschlicherweise als Land eingeordneten Pixelklasse, die stets einen Anteil von ~0,2 % aufweist (Tab. 3). Diese Pixel befinden

sich v.a. in der Nähe des Rheinufers (Abb. 5) und sie werden wahrscheinlich durch Rückstreueffekte ausgelöst, die z.B. Bühnen, Stege und andere Bauwerke verursachen.

Somit wird die Fehlklassifizierung beider Klassen durch die Verwendung aller Zusatzdaten von 9,73 % (Summe falsch klassifizierter Land- und Wasser-Pixel aus Tab. 3) auf 0,64 % (Summe falsch klassifizierter Land- und Wasser-Pixel aus Tab. 2) reduziert.

Tab. 1: Konfusionsmatrix für die Wassermaske, die aus S-1 SLC IW VV Daten des 17.11.2018 und Verwendung aller Zusatzdaten (Fahrrinne, Landbedeckung, HAND-Index) abgeleitet wurde. Angegeben ist die jeweilige Pixelanzahl.

		Validierungsdaten [Pixel]		
		Land	Wasser	Summe
Klassifikation [Pixel]	Land	<b>1.491.101</b>	3.455	1.494.556
	Wasser	6.522	<b>45.410</b>	51.932
	Summe	1.497.623	48.865	<b>1.546.488</b>

Tab. 2: Konfusionsmatrix für die Wassermaske, die aus S-1 SLC IW VV Daten des 17.11.2018 und Verwendung aller Zusatzdaten (Fahrrinne, Landbedeckung, HAND-Index) abgeleitet wurde. Angegeben ist der prozentuale Anteil bezogen auf die Gesamtpixelzahl.

		Validierungsdaten [%]		
		Land	Wasser	Summe
Klassifikation [%]	Land	<b>96,42</b>	0,22	96,64
	Wasser	0,42	<b>2,94</b>	3,36
	Summe	96,84	3,16	<b>100,00</b>

Die Richtigkeit einer entsprechenden Objektklasse aus Herstellersicht (*producer accuracy*, PA) wird durch Division der Zahl korrekt klassifizierter Pixel durch die Anzahl der klassifizierten Pixel der entsprechenden Klasse der Validierungsdaten (hier Wassermaske aus den DOPs) bestimmt. Die PA spiegelt somit die Wahrscheinlichkeit wieder, mit der das entsprechende Pixel korrekt zur jeweiligen Klasse zugeordnet wurde. Analog ergibt die Anzahl korrekt klassifizierter Pixel der jeweiligen Klasse, dividiert durch die Pixel-Summe der jeweiligen Klasse, die Richtigkeit aus Nutzersicht (*user accuracy*, UA, Tab. 3). Dieses Qualitätsmaß gibt somit die Wahrscheinlichkeit an, mit der ein Pixelwert einer bestimmten Klasse auch tatsächlich dieser Klasse angehört.



Tab. 3: Auswertungsmatrix für die Wassermasken, die aus S-1 SLC IW VV Daten des 17.01.2018 abgeleitet wurden. Angegeben sind die prozentualen Anteile an der Gesamtpixelzahl der klassifizierten Pixel im Vergleich zu den entsprechenden Wasser-/Land-Pixel des DOP vom 16.11.2018. UA (user accuracy) und PA (producer accuracy) beziehen sich auf den Vergleich mit den entsprechenden Wasser-/Land-Pixel des DOP vom 16.11.2018.

Klassifikation im DOP	Land				Wasser			
Pixelanzahl	1.497.623				48.865			
[%]	96,84%				3,16%			
Klassifizierung [%]	korrekt als Land klassifiziert	falsch als Land klassifiziert	UA	PA	korrekt als Wasser klassifiziert	falsch als Wasser klassifiziert	UA	PA
<b>S1</b>	87,34	0,23	99,74	90,19	2,93	9,50	23,59	92,87
<b>S1 + Fahrrinne</b>	87,34	0,22	99,75	90,19	2,94	9,50	23,61	92,95
<b>S1 + Landbedeckung</b>	95,27	0,23	99,76	98,38	2,93	1,57	65,18	92,85
<b>S1 + HAND-Index</b>	95,67	0,23	99,77	98,80	2,96	1,17	71,56	92,87
<b>S1 + alle Zusatzdaten</b>	<b>96,41</b>	<b>0,22</b>	<b>99,77</b>	<b>99,56</b>	<b>2,94</b>	<b>0,42</b>	<b>87,44</b>	<b>92,93</b>

Durch den Einsatz aller Zusatzdaten wurden die UA der Klasse Wasser um circa 64 % gesteigert, die Klasse Land hat nur ~9 % Verbesserung zu verzeichnen (Tab. 3). Die starke Verbesserung der UA-Werte der Wasserklassifikation hängt mit der deutlichen Senkung falsch als Wasser klassifizierter Pixel zusammen. Bei der PA hingegen konnte die Richtigkeit in der Wasser-Klasse nicht wesentlich gesteigert werden (nur geringer Zuwachs korrekt klassifizierter Wasser-Pixel). Wogegen bei der Land-Klasse eine Steigerung von ~9 % erreicht wird (Tab. 3). Sowohl visuelle, als auch quantifizierende Validierungsmaßnahmen zeigen, dass Verbesserungen durch Zusatzdaten größtenteils über Landbereichen erfolgen und die meisten falsch als Wasser klassifizierten Pixel zu Land-Pixel korrigiert werden, was in der zunehmenden UA der Wasser-Klasse deutlich wird.

## 6 Fazit & Ausblick

Das Anwendungsbeispiels Niedrigwasser am Mittelrhein 2018 zeigt, dass ausschließlich aus S-1 Radardaten extrahierte Wasserflächen einige Fehlklassifizierungen nach sich ziehen. Da die Problematiken messtechnisch bedingt sind (z.B. Abschattungsbereiche, Überlagerungseffekte), wurden zur Reduktion der Fehlklassifizierungen GeoBasisdaten wie DGM5, Landnutzungsdaten, sowie die Lage der Fahrrinne verwendet. Die deutlichsten Verbesserungen der Wasser-Land-Ableitung konnten durch Integration von Höhen- und Landbedeckungsinformationen erreicht werden, die ohne Zusatzdaten fälschlicherweise als Wasser klassifiziert wurden. Insgesamt wurde eine Richtigkeit aus Nutzersicht (UA) von 87,44 % für Wasserflächen bzw. 99,77 % für Landflächen erreicht. Die verbleibenden Abweichungen befinden sich vor allem in direkter Nähe zum Fluss und

werden vermutlich durch Abschattungseffekte und die geringe Auflösung der S-1 Daten hervorgerufen. Für detaillierte Betrachtungen des trockenfallenden Bereichs sind diese Differenzen jedoch relevant, so dass die Ableitungen aktuell noch nicht den Anforderungen für ein Monitoring von Extremereignissen, ein Veränderungsmonitoring oder zur Überwachung der Schiffbarkeit genügen. Im weiteren Projektverlauf wird darum zusätzlich das hochaufgelöste Digitale Geländemodell des Wasserlaufs (DGM-W) integriert, geprüft wieweit sich dadurch die noch vorhandenen Unsicherheiten in der Nähe der WLG reduzieren und die Anwendbarkeit detailliert herausgearbeitet.

## 7 Danksagung

Wir bedanken uns für die Förderung des Projektes „Sat-Land-Fluss“ (Förderkennzeichen 50EW2015) in der Förderrichtlinie „Entwicklung und Implementierungsvorbereitung von Copernicus Diensten für den öffentlichen Bedarf in Deutschland“, durch das Bundesministerium für Digitales und Verkehr (BMDV) und den Projektträger des Deutschen Zentrums für Luft und Raumfahrt (DLR).

## 8 Literaturverzeichnis

- BBK, 2023: Copernicus Dienst für Katastrophen- und Krisenmanagement. [https://www.bbk.bund.de/DE/Themen/Krisenmanagement/Lagebild/Fernerkundung/Copernicus/copernicus\\_node.html](https://www.bbk.bund.de/DE/Themen/Krisenmanagement/Lagebild/Fernerkundung/Copernicus/copernicus_node.html), letzter Zugriff 04.01.2023.
- EODC, 2023: Global Flood Monitoring Portal. <https://portal.gfm.eodc.eu/>, letzter Zugriff 05.01.2023.
- ESA, 2022: Geographical Coverage. <https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-1/satellite-description/geographical-coverage>, letzter Zugriff am 29.12.2022.
- ESA, 2023a: About WorldCover. <https://esa-worldcover.org/en/about/about>, letzter Zugriff 11.01.2022.
- HUANG, W., DEVRIES, B., HUANG, C., LANG, M.W., JONES, J.W., CREED, I.F. & CARROLL, M.L., 2018: Automated Extraction of Surface Water Extent from Sentinel-1 Data. *Remote Sensing*, **10**, 797, <https://doi.org/10.3390/rs10050797>.
- JENEROWICZ, J., ORYCH, A., SIOK, K. & SMIAROWSKI, M., 2019: Application of Sentinel-1 and Sentinel-2 data to conduct reconnaissance analyses. *SPIE 11160*, *Electro-Optical Remote Sensing*, **13**, 111600L, <https://doi.org/10.1117/12.2533319>.
- KAPLAN, G., LIOR F., LUKYANOV, V., MANIVASAGAM, V. S., TANNY, J. & ROZENSTEIN, O., 2021: Normalizing the Local Incidence Angle in Sentinel-1 Imagery to Improve Leaf Area Index, Vegetation Height, and Crop Coefficient Estimations. *Land*, **10**(7), 680, <https://doi.org/10.3390/land10070680>.
- KITTLER, J. & ILLINGWORTH, J., 1985: On threshold selection using clustering criteria. *IEEE Transactions on Systems, Man, and Cybernetics*, *SMC-15* (5), 652-655. <https://doi.org/10.1109/TSMC.1985.6313443>.
- KUMAR, D., 2021: Urban objects detection from C-band synthetic aperture radar (SAR) satellite images through simulating filter properties. *Scientific Reports*, **11**, 6241.

<https://doi.org/10.1038/s41598-021-85121-9>.

- LIAO, H.-Y. & WEN, T.-H., 2020: Extracting urban water bodies from high-resolution radar images: Measuring the urban surface morphology to control for radar's double-bounce effect. *International Journal of Applied Earth Observation and Geoinformation*, **85**, 102003, 1-13  
<https://doi.org/10.1016/j.jag.2019.102003>.
- MARTINIS, S., KERSTEN, J. & TWELE, A., 2015: A fully automated TerraSAR-X based flood service. *ISPRS Journal of Photogrammetry and Remote Sensing*, **104**, 203-212,  
<https://doi.org/10.3390/rs10040583>.
- MARTINIS, S., PLANK, S. & ČWIK, K., 2018: The Use of Sentinel-1 Time-Series Data to Improve Flood Monitoring in Arid Areas. *Remote Sensing*, **10**(583), 1-13,  
<https://doi.org/10.3390/rs10040583>.
- NOBRE, A. D., CUARTAS, L. A., HODNETT, M., RENNÓ, C. D., RODRIGUES, G., SILVEIRA, A., WATERLOO, M. & SALESKA, S., 2011: Height Above the Nearest Drainage – a hydrologically relevant new terrain model. *Journal of Hydrology*, **404**, 13-29,  
<https://doi.org/10.1016/j.jhydrol.2011.03.051>.
- OLTHOF, I. & TOLSZCZUK-LECLERC, S., 2018: Comparing Landsat and RADARSAT for Current and Historical Dynamic Flood Mapping, *Remote Sensing*, **10**(5), 780, 1-19,  
<https://doi.org/10.3390/rs10050780>.

# Applied Space-borne Remote Sensing to Identify Mass Movements and the Exemplary Modelling of Potentially Catastrophic Failures in the Bhagirathi Area, India

GWENDOLYN DASSER<sup>1,2</sup>, JESSICA MUNCH<sup>1,2</sup>, YVES BÜHLER<sup>1,2</sup>,  
PERRY BARTELT<sup>1,2</sup> & ANDREA MANCONI<sup>1,2</sup>

*Abstract: Extreme rainfall intensities, degrading permafrost and glacier melt in combination with ongoing construction of hydropower plants and settlement growth are leading to an ever-increasing amount of mass movements in alpine terrain. Geologically young mountain ranges such as the Indian Himalaya are experiencing severe consequences due to the combination of climatic changes and anthropogenic expansion, leading to higher risk for human lives and infrastructures. We present an approach to generate hazard indication maps of mass movements in high alpine areas using a combination of remote sensing and numerical simulations. We use satellite radar interferometry from Sentinel-1 and geomorphological feature classification based on optical satellite data. We show two exemplary case studies and model one of these to assess its hazard potential using the newly developed RAMMS::ROCKICE module.*

## 1 Introduction

Systematic mapping, monitoring, and modelling of slope instabilities are essential tools for an effective hazard assessment, risk management and disaster response. Extreme topographic reliefs such as the Indian Himalayas are especially affected by current climatic changes such as weather extremes, glacier retreat or permafrost thawing (DAI et al., 2023). More and more large and potentially catastrophic mass movements are observed. Especially in the state of Uttarakhand, India, past events such as the rock and ice avalanche in Chamoli in 2021 (SHUGAR et al. 2021; VERMA et al. 2021; PANDLEY et al. 2022), have led to an increasing demand to identify and forecast potential geohazard scenarios. Several hydropower projects as well as new roads are currently under construction or in the planning phase in this area, increasing the risk to infrastructure and human lives (DEVARA et al. 2021), while also leading to further slope instabilities in the area (SATI et al. 2020).

Accurate measurements may allow detecting changes and/or precursors of such potentially catastrophic evolutions of mass movements. The Himalayas are a remote and challenging terrain, thereby strongly restricting the possibilities for in-situ measurements and especially systematic monitoring (DINI et al. 2019). Space-borne imagery allows a cost-effective possibility to provide regular information on such areas (STROZZI et al. 2004), also allowing to cover entire states, thereby complementing local measurements and observations (LIU et al. 2022).

---

<sup>1</sup> WSL Institute for Snow and Avalanche Research SLF, Alpine Remote Sensing, Flüelastrasse 111, CH-7260 Davos Dorf, Switzerland, E-Mail: [gwendolyn.dasser, andrea.manconi, yves.buehler, jessica.munch, perry.bartelt]@slf.ch

<sup>2</sup> Climate Change, Extremes and Natural Hazards in Alpine Regions Research Center CERC, Flüelastrasse 111, CH-7260 Davos Dorf, Switzerland

Mass movements such as landslides can be detected and classified at an early stage using a combination of interferometric data and optical imagery (JONES et al. 2021). Ground displacements can be measured at sub-centimetre scale using differential interferometric synthetic aperture radar (DInSAR; PASQUALI et al. 2014), thereby also allowing a detection of movements that might not yet be visible in the field (ZHOU et al. 2022). Such small-scale slope displacements may evolve over time and in some cases precede ultimate failures (MANCONI 2021). Optical remote sensing enables to retrace origin of movement to corresponding geomorphological processes. During cloud-free conditions, the usage of optical sensors also allows to cover areas that might be invisible to radar due to strong geometric effects in alpine areas. Hence, exploiting the combination allows to identify a higher amount and more diverse potential geohazards.

Current research of hazards in the state of Uttarakhand has been focused on post-event analysis. The rock and ice avalanche in Chamoli has been thoroughly examined by making use of satellite imagery, seismic records as well as eyewitness videos and numerical models (SHUGAR et al. 2021). Further research was conducted on the effect of vegetation loss to flash flood in the same case scenario (VERMA et al. 2021). Extensive hazard assessment using DInSAR is currently conducted over the area of Josimath, Uttarakhand, for diverse signs of ongoing ground movement have been found within the village (see DISCUSS TERRADUE 2023).

In this work, we present a workflow for large scale hazard analysis based on space-borne remote sensing for high alpine terrain of the Indian Himalaya. We provide a first attempt on systematic and spatially continuous assessment of unstable areas for the Bhagirathi valley using DInSAR. Using optical images, data gaps in radar-based images can be covered, and the identified mass movements specifically analysed to detect signs of activity and classify the geomorphological features. A cost-effective approach to further detect potential hanging glacier failures as seen in the event of Chamoli 2021 is also proposed. Two areas where catastrophic events may possibly develop in the future are then discussed in detail, and for one case the hazard potential is modelled using the Rock-Ice module of the Rapid Mass Movement Simulation (RAMMS) tool.

## 2 Methods

### 2.1 Research Site

The Bhagirathi valley, located in the state of Uttarakhand in northern India, is of particular interest due to the amount of infrastructure and potentially exposed population in case of catastrophic events. The valley contains multiple smaller clusters of infrastructure such as the settlements of Bagori, Harshil and Naga, which are situated on the sedimented areas in river deltas and several hydropower projects in different state of planning and construction. The area experiences heavy rainfall during the Indian Summer Monsoon (BARNAND et al. 2004). Geomorphological analysis of the Bhagirathi catchment indicates periglacial processes present above 2000 m, while it being the driving factor above 4000 m elevation (SATI et al. 2020). The fluvial erosion, the location in periglacial zone and heavy and focussed rainfall lead to the valley being especially prone to mass movement activity (BARNARD et al. 2004) and hence being in the physiographic transition zone.

Fig. 1 indicates the geolocation of the research site featuring the extent of the area processed with DInSAR based on corresponding S1 tracks. Two locations are indicated, which are examples for identified hazards.

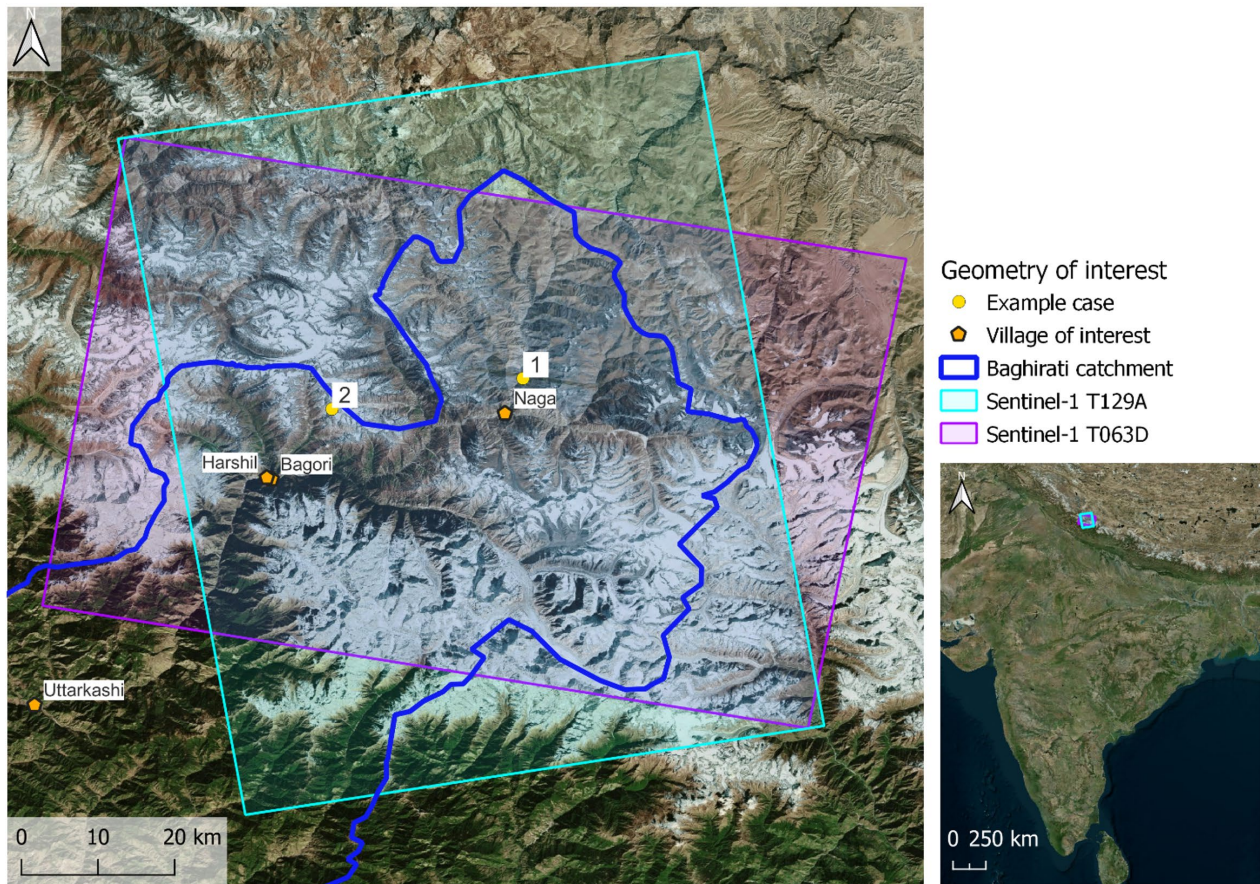


Fig. 1: Map featuring the locations of the two example cases and infrastructure of interest in the Bhagirathi valley, India. Indicated are used footprints from Sentinel-1 in ascending and descending node. Base map by BingAreal (Image acquisition date unknown).

## 2.2 Data and Preprocessing

Throughout this project we use a spectrum of remote sensing data and exploit combinations to detect, map and identify natural hazards. The combination of multiple data sources allows to capture various types of potentially catastrophic failures and enables to detect hazards which might be occluded in one stand-alone data analysis. We include information retrieved from Sentinel-1, Sentinel-2, PLANET, Bing Areal, Google Earth, digital elevation models (DEM), as well as a publicly available permafrost layer (GRUBER 2012) and glacier inventory (GLIMS CONSORTIUM 2005).

We process and analyse satellite radar datasets acquired from the ESA Sentinel-1 (S1) mission in the period 2018-2021 for the footprints indicated in Fig. 1. We selected one ascending (T129) and one descending (T063) track to gain information in two line-of-sight directions (LOS), thereby minimizing the number of shadowed areas. The SAR dataset is composed of 117 images in ascending and 103 images in descending orbit. We use standard DInSAR to first identify and then classify areas affected by potential instabilities. Interferograms were computed by selecting image-pairs based on perpendicular baseline below 3 m and temporal baseline of below 1 year. This resulted in a dataset featuring interferometric baselines between 12 to 300 days. From these, interferograms were manually discharged, where visual analysis indicated coherence within the Bhagirathi valley to not be feasible. Information on the interferograms is summarized in table 1.

We made use of a combination of base map analysis by Google Earth, Bing areal and PLANET data. We included PLANET data, even though the spatial resolution is coarser, to analyse deformation changes over specific timeframes. Selection of dates of PLANET acquisition is indicated in table 1 and was done by searching for the most feasible trade-off between snow, cloud and areal coverage while covering at least all summer seasons.

Tab. 1: Temporal windows of S1 image-pairs per track and acquisition date of selected PLANET data

Track 129 ascending		Track 063 descending		PLANET
Primary image	Secondary image	Primary image	Secondary image	
30.08.2018	26.06.2019	22.01.2018	10.05.2018	22.01.2018
30.10.2020	09.07.2021	22.05.2018	21.07.2018	22.05.2018
23.11.2020	23.03.2021	16.07.2019	21.08.2019	04.08.2018
05.12.2020	17.12.2020	29.04.2020	04.06.2020	23.07.2019
05.12.2020	03.02.2021	26.10.2020	18.01.2021	07.09.2019
17.12.2020	03.02.2021	26.10.2020	05.07.2021	12.06.2020
23.03.2021	14.08.2021	18.01.2021	05.07.2021	19.10.2020
				27.02.2021
				01.07.2021

In the framework of this project, we used a publicly available glacier inventory (GLIMS CONSORTIUM 2005) to identify possible hanging glaciers by intersecting the dataset with a slope mask of above 25 degrees. This terrain slope was calculated from the 30 m Copernicus digital elevation model (DEM). To better identify features created through processes in connection to permafrost, the freely available global permafrost zonation indication map (GRUBER 2012) was consulted.

### 2.3 RAMMS::ROCKICE simulations

RAMMS::RockIce is a thermo-mechanical module from the RAMMS software (CHRISTEN et al., 2010, BARTELT et al., 2018) that is specifically designed to model gravitational flows made of rock/ice/water/snow or any combination of these. The module is based on the depth-averaged conservation of mass, momentum, and energy for all the phases involved in the flow and allows for interactions in between the phases, such as heat transfer between the different materials as the temperature of each material is computed separately during a simulation. This means that the ice can melt and turn into water, increasing the water content of the flow and potentially the runout distances (SCHNEIDER et al. 2011). The module also allows for ground material entrainment following (FRANK et al. 2015), which again can impact the distances travelled by the flow.

In this work, we are interested in different setups for gravitational movements, involving different kinds of materials. Firstly, we are investigating a deep seated landslide, and secondly an ice avalanche starting from a hanging glacier. The remote sensing methodology, the numerical modelling as well as the main results for one of these cases will be developed in part 3.3.

### 3 Identified geomorphological hazards and their modelling

#### 3.1 Geohazard assessment in the greater Bhagirathi area

The examination of interferograms fringe patterns across diverse baselines, allows a detection and mapping of mass movements featuring different velocities. By analysing images acquired using ascending and descending tracks, movement in west-east and east-west, respectively, can be detected. By visual comparison of fringe patterns to optical imagery geomorphological features were identified. In this study, these features mainly included landslides, (rock) glaciers and fluvial sediment relocations. Large deep-seated landslides were often detectable across temporal baselines; however, the smaller temporal windows allowed to differentiate and detect also faster, secondary landslides.

Above-average mass movement activity in two side valleys neighbouring the Bhagirathi valley were identified, including two large instabilities with a velocity of more than 30 cm/year in descending LOS). Around twenty more areas show velocities of more than 15 cm/year in more than one interferogram. In the following, we selected two example cases based on the potential impact on infrastructure and corresponding hazard potential.

#### 3.2 Example case 1: Deep seated landslide failure

We discuss here an example of a large, deep-seated landslide, which already resulted in visible displacement in optical imagery. At the bottom of the landslide, a river connects the moving area to the village of Naga, further downstream. The river course has been altered due to ground movement and a crest is visible (see Fig. 2). A road follows this river, also being built upon the landslide area. The road crosses the river twice via bridges on the way downstream to the city of Naga. DInSAR data indicates the landslide to be a complex moving system featuring different layers and velocities. According to the permafrost layer created in 2018, permafrost presence is possible but uncertain. With a global decrease in permafrost, the slope instability might therefore be caused or at least fostered by permafrost degradation processes. The interferograms indicate three different movement zones (see Fig. 4).

In the framework of this report, we analyse one scenario of a sudden release of a part of this landslide, beginning at the indicated visible crest (see Fig. 2). Depth of this moving ground mass has been estimated via the area of the landslide and the equation and variables provided in JONES (2021) to be around 20 m.

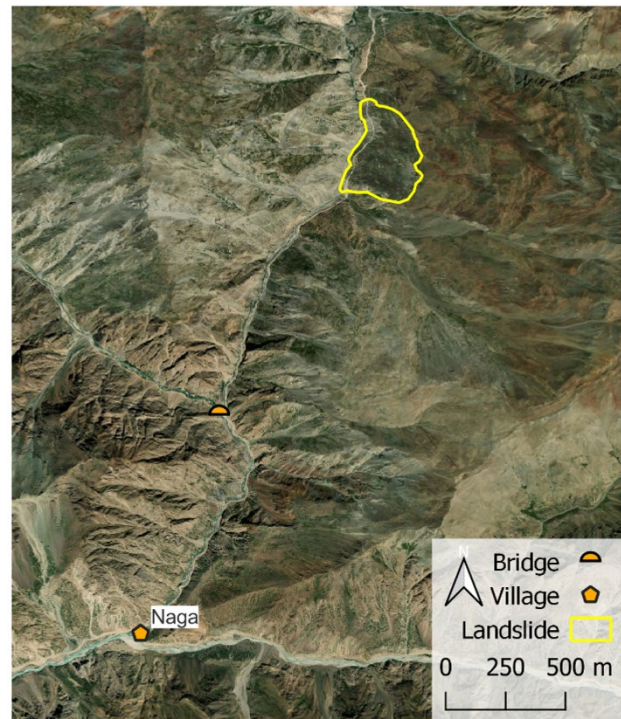


Fig. 2: Situation of landslide case study in relation to a bridge, the village of Naga on Bing Aerial base maps (Image acquisition date unknown).



Tab. 2: Temporal windows of interferograms (primary image named first), the corresponding temporal baseline, the detected range of fringes and calculated surface velocity when upscaled to a year

Temporal window interferograms [dd.mm.yyyy – dd.mm.yyyy]	Temporal baseline [days]	Surface velocity away from LOS [cm/year]
22.01.2018 – 10.05.2018	108	2.4 to 4.7
22.05.2018 – 21.07.2018	60	8.5 to 11.3
16.07.2019 – 21.08.2019	36	14.2 to 28.4
29.04.2020 – 04.06.2020	36	14.2 to 28.4
26.10.2020 – 18.01.2021	92	5.6 to 11.1
26.10.2020 – 05.07.2021	252	4.1 to 8.1
18.01.2021 – 05.07.2021	168	6.1 to 12.2

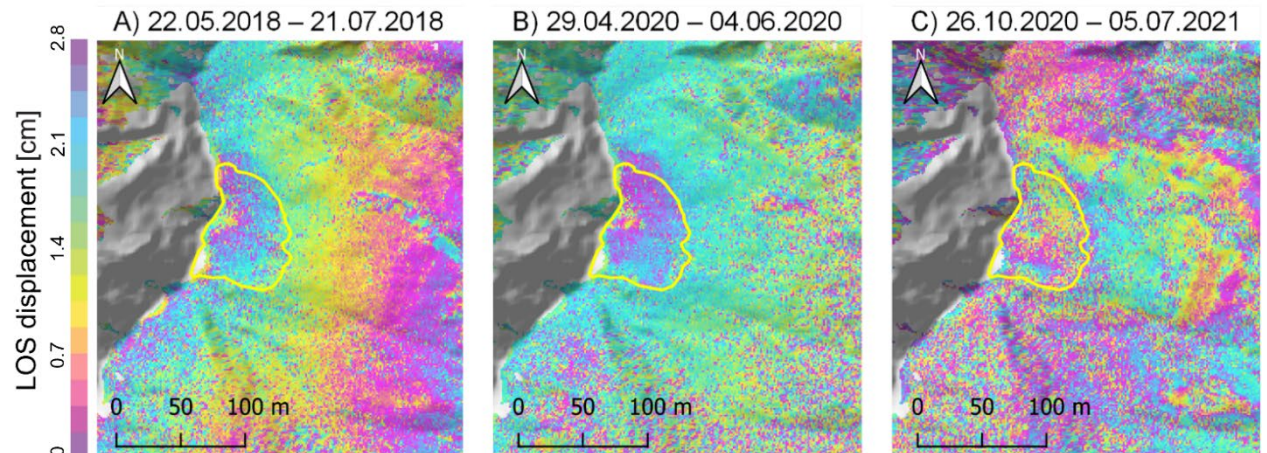
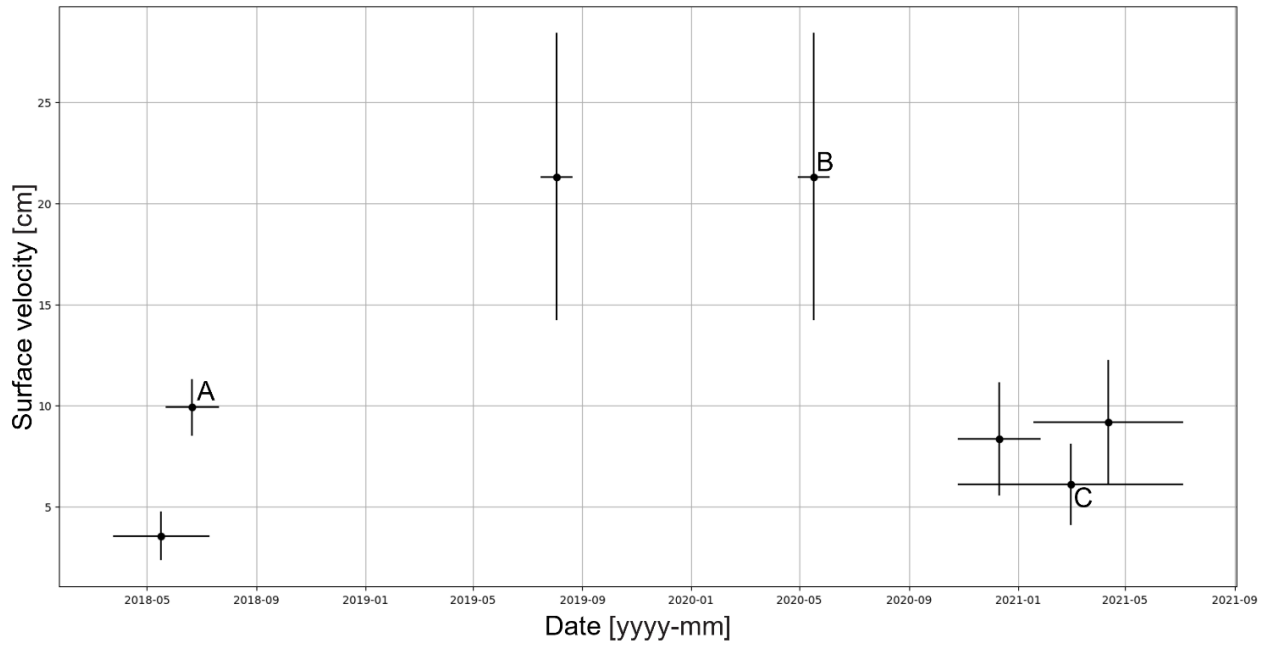


Fig. 3: Magnitude of estimated surface velocities of the slope deformation across interferograms featuring different time windows in descending orbit and three exemplary interferograms

The results of the interferometric analysis of images are summarized in Tab. 2 and illustrated in Fig. 3. Only interferograms in descending orbit were feasible for the detection of this movement

due to the orientation in relation to the LOS. Each interferogram indicates an occurring phase difference (see examples in Fig. 3).

The interferogram of May 2018 to July 2018 shown in Fig. 3.A) indicates movement, also visible is large-scale atmospheric phase delay, which is correlated with topography. Between April and July 2020 (Fig. 3.B) a secondary, faster landslide within the delineated, already moving area, is apparent. The phase difference between October 2020 and July 2021 (Fig. 3.C) indicates a further complexity of the moving mass, as a third, even larger area is also – however, more slowly – moving.

Derived surface velocities are constantly above background noise. Highest surface velocities are found in interferograms of smallest temporal baseline. This indicates sharp increase of surface displacements, which might be due to snowmelt in spring, or heavy rainfall during the monsoon season. Longer temporal windows average the movement over time, resulting in lower velocity values, as to be expected (see Fig. 3).

These findings show clear signs of surface deformation over time. Slopes characterised by such ongoing movements are instable and can be triggered in case of extreme rainfall intensity/duration or an earthquake, leading to a sudden release and a potentially catastrophic outcome. Together with the vicinity of the landslide to infrastructure such as the bridge indicated in Fig. 2 and the village of Naga further downstream, this landslide poses threat to human lives in such a case. Currently, we are in the modelling phase of potential scenarios and first results indicate a strong dependency of atmospheric precipitation on the landslides hazard potential.

### 3.3 Example case 2: Hanging glacier failure

The rock and ice avalanche of Chamoli has shown hanging glaciers to pose risk in case of sudden release (SHUGAR et al. 2021). However, glacier movement is near to impossible to detect on interferometric data, as the highly dynamic environment is changing too rapidly for phase analysis to be effective. We therefore used the global glacier inventory. Using an intersection between slope above 25 degrees and the glacier inventory, a hanging glacier library was created for the area of interest. This library was imported to Google Earth and the potential hazard for infrastructure manually assessed for each case. The glacier with the highest hazard potential was selected and further investigated. Based on optical images, the most likely failure scenario was delineated, following already visible crevasses in the glacier (see Fig. 4). Glacier depth of 50 m was estimated by averaging glacier front measurements at the possible failure location in Google Earth.

The release zone was defined as a block of pure ice, with a depth of 50 m and a total volume of  $\sim 250\,000\text{ m}^3$ . The release area was delineated between two sets of converging cracks in the glacier as found in optical imagery (see Fig. 4).

Two different scenarios of this hanging glacier release were modelled using RAMMS::ROCKICE. The two scenarios diverge only in the presence of atmospheric snow cover: We defined a snow-free and a snow-rich scenario for the track of the ice avalanche. In the latter, 7 m of snow cover depth was added to the model, starting at release zone with a reduction of 0.03 m per 100 m elevation decrease.

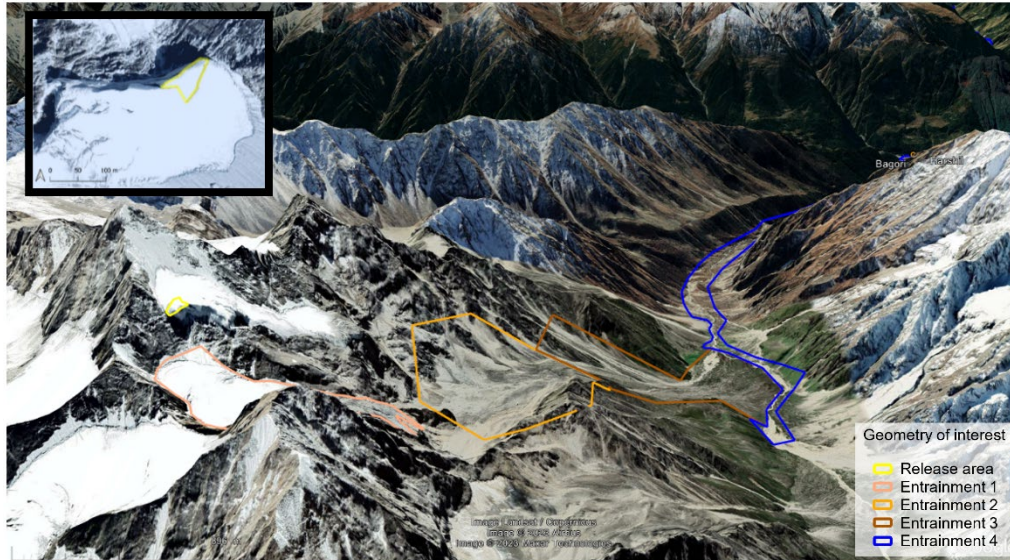


Fig. 4: Situation of potential hanging glacier failure in relation to the villages of Bagori and Harshil and a top view on the delineated failure part (image from 18/8/2022, Google Earth). Indicated are the applied entrainment zones (parameters described in part 3.3)

The potential runout path in case of failure covers diverse terrains. These range from glacier over pre-glacial sediments and a sediment cone to the river-basin. To model this range of terrain types, we defined four different entrainment zones that are shown in Fig. 4 and the parameters selected are summarized in table 3. The release as well as each of the entrainment layers were set at a temperature of 0 °C in both scenarios. The simulations were ran for 3000 s or until the momentum of the flow was less than 3% of its initial value.

Table 3: Parameters set for each entrainment zone in both case scenarios in case of glacial collapse.

Entrainment [No.]	Material composition [rock%, ice%, water%]	Density [kg/m <sup>-3</sup> ]	Erosion rate [m/s <sup>-3</sup> ]	Potential erosion depth [m]	Critical shear stress [kPa]	Maximal erosion depth [m]
1	0, 95, 5	1000	0.013	0.100	1	10
2	60, 30, 10	2000	0.013	0.100	1	10
3	75, 15, 10	2400	0.05	0.1	0.5	10
4	80, 0, 20	2500	0.05	0.2	0.5	10

In the snow-free glacier collapse scenario (Fig. 5.A), areas affected by the released material is minimal. The short runout also results in a very limited effect of entrained material. In the snow-rich scenario however (Fig. 5.B), the snow cover results in more material present in the release at model initiation, but also in the entrainment zones. This leads to a longer runout due to more material being entrained and also a potential to produce a higher amount of meltwater as both snow and ice as susceptible to melt. High water content in mass movement increases the mobility of the materials and thus their potential for destruction at further distances (SCHNEIDER et al., 2011). This is due to water reducing the ground friction as reported in the case of Chamoli 2021 (SHUGAR et al. 2021). The resulting higher velocity and longer runout of the moving mass results in more mass being transported downstream. In this scenario, calamities and economic losses are to be expected.

The current snow-rich scenario setup and corresponding parameters used, indicates a complete destruction of the villages of Harshil and Bagori.

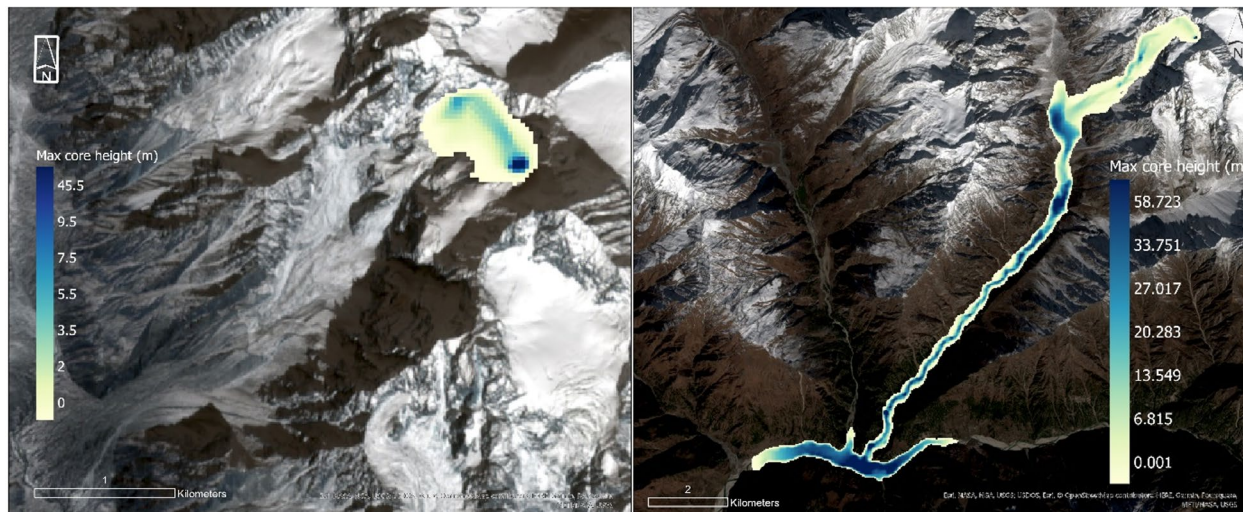


Fig. 5: RAMMS simulations of the ice avalanche for two scenarios. The maximal core height in meters is shown for the hanging glacier collapse for a snow-free (left, A) and snow-rich (right, B) scenario including entrainment. Background map is the Sentinel-2 RGB-image acquired on 11.04.2022.

### 3.4 Challenges

A visual analysis of the entire Bhagirathi valley as performed here is time consuming. Hitherto, no automatization process is in place to make such a hazard indication map generation more efficient. Currently we need to filter areas to efficiently map ground movement with highest impact potential. In the framework of this project, we focus on higher elevations zones, where glaciers and permafrost are key components in the landscape. The event at Chamoli has shown that ice can lead to an even more catastrophic result when melted by frictional heat (SHUGAR et al. 2021). We thus include in our geohazard assessment also processes of corresponding nature. The consultation of external datasets for e.g. glacier localisation might result in further uncertainties in the processing chain. For being a highly dynamic environment, archives based on historic data might not be up to date or datasets might feature data gaps we overlooked. This could result in missing out potentially failing geohazards. As a result, we do not capture all potentially hazardous events and are still limited in spatial coverage due to high processing and working costs.

High alpine terrain as the one present includes a range of challenges for remote sensing approaches. DInSAR images for example suffer from geometric effects such as shadowing, foreshortening and layover (MANCONI 2021). Atmospheric phase delay is further amplified in terrain featuring high topographic gradients (CIGNETTI et al. 2016), rendering the technique partly inapplicable in such areas. Vegetational activity and heavy rainfall events such as the Monsoon result in decorrelation of image pairs, further restricting the usage of phase information. Fast and complex ground movement can further lead to phase ambiguity. Geometric restrictions of spaceborne radar also makes it impossible to capture movement in unsuitable movement directions (COLESANTI & WASOWSKI 2006). DInSAR datasets therefore allow to detect some movement in alpine terrain, however it does not cover all hazard types or movement directions, nor does it provide gap free information (MANCONI 2021).

When performing the modelling, multiple assumptions had to be made, mostly regarding the composition of the entrainment layers. In this study, we estimated such parameters based on optical satellite image interpretation, as well as on experience gained from modelling other mass movements containing rock/ice/water/snow or a mixture of these, such as the post-event modelling of Chamoli. As water content in the flow has a major impact on the runout distances of the modelled flows, the results are highly dependent on the composition and physical properties of the release and entrainment zones; and are only valid within the framework of the assumptions that were made. The models were run for two scenarios, one featuring snow coverage and one free of snow. Further research should include also other effects of seasonality and weather extremes such as the Monsoon, as for heavy rainfall results in a higher water content in all influenced stages of entrainment, leading to larger runouts and changes in fluid dynamics. Larger amounts of water can further trigger potential (additional) landslides in the flow.

By highlighting the influence of the composition of the release zone and entrainment layers, we want to further insist on the importance of field observations and measurements. These are required to better understand geomorphological features present, the compounds and extents of possible entrainment zones included in the process and climatic and weather data.

## 4 Conclusions & Outlook

Using a combination of various remote sensing techniques, we provide a first assessment on potential hazards present in the Bhagirathi area. By exploiting high resolution optical imagery such as provided by PLANET and Google Earth, we classified ground deformation information mapped using DInSAR according to geomorphological processes. Our results indicate deep seated landslides moving up to 30 cm/year, accelerating rock glaciers as well as highly dynamic hanging glaciers to be potential geohazards present in the Bhagirathi area.

Based on two different case studies, we discussed challenges and opportunities of the different datasets and showed that for a complete hazard indication map a combination of datasets is invaluable. Using DInSAR technology in combination with optical image analysis, a complex landslide was detected, which features movement already over several years and with estimated velocities of up to 14-28 cm/year. Using a combination of glacier inventories, optical imagery, and slope information, we identified potential hazard through hanging glacier collapses.

With the usage of the new Rock-Ice avalanche module in RAMMS, we delineated the potential hazard in one case study. It indicated strong potential for calamities and destruction of downstream infrastructure especially in scenarios containing high amounts of snow and ice. This further highlights the effect of atmospheric precipitation on the hazard potential of such scenarios.

We pursue this project further by setting up a workflow for entire Uttarakhand and providing an overview on mass movement activity in the area with potential catastrophic consequences. This enables local researchers to focus on areas of potential hazards. Specific field work can then be conducted to minimize uncertainties in hazard modelling by increasing the understanding of the local environment. In the framework of planned workshops, we also transfer knowledge on the processing workflow to local researchers in India, allowing to create a dynamic mass movement inventory of the state of Uttarakhand.

## 5 Acknowledgment

This work is supported by the Swiss international cooperation and humanitarian aid under the SDC project “*Thermomechanical modelling of rock-ice avalanches for large scale hazard indication mapping in Uttarakhand Province*” implemented by the contractor WSL/SLF.

## 6 Literature

- BARTELT, P., CHRISTEN, M., BÜHLER, Y. & BUSER, O., 2018: Thermomechanical modelling of rock avalanches with debris, ice and snow entrainment. Numerical methods in geotechnical engineering IX, Porto, PT, 1047-1054.
- BARNAND, P.L., OWEN, L.A. & FINKEL, R.C., 2004: Style and timing of glacial and paraglacial sedimentation in a monsoon-influenced high Himalayan environment, the upper Bhagirathi Valley, Garhwal Himalaya. *Sedimentary Geology*, **165**, 199-221.
- CHRISTEN, M., KOWALSKI, J. & BARTELT, P., 2010. RAMMS: Numerical simulation of dense snow avalanches in three-dimensional terrain. *Cold Regions Science and Technology*, **63**, 1-14.
- CIGNETTI, M., MANCONI, A., MANUNTA, M., GIORDAN, D., DE LUCA, C., ALLASIA, P. & ADRIZZIONE, F., 2016: Taking Advantage of the ESA G-POD Service to Study Ground Deformation Processes in High Mountain Areas: A Valle d’Aosta Case Study, Northern Italy. *Remote Sensing*, **8**(852).
- COLESANTI, C. & WASOWSKI, J., 2006: Investigating landslides with space-borne Synthetic Aperture Radar (SAR) interferometry. *Engineering Geology*, **88**, 173-199.
- DAI, C., LI, W., LU, H. & ZHANG, S., 2023: Landslide Hazard Assessment Method Considering the Deformation Factor: A Case Study of Zhouqu, Gansu Province, Northwest China. *Remote Sensing*, **15**(596).
- DEVARA, M., TIWARI, A. & DWIVEDI, R., 2021: Landslide susceptibility mapping using MT-InSAR and AHP enabled GIS-based multi-criteria decision analysis. *Geomatics, Natural Hazards and Risk*, **12**(1), 675-693.
- DINI, B., DAOUT, S., MANCONI, A. & LOEW, S., 2019: Classification of slope processes based on multitemporal DInSAR analyses in the Himalaya of NW Bhutan. *Remote Sensing of Environment*, **233**(111408).
- DISCUSS TERRADUE, 2023: Results of advances InSAR services (Sentinel-1) indicate that the town of Joshimath (Northern India) is sliding. <https://discuss.terradue.com/t/results-of-advanced-insar-services-sentinel-1-indicate-that-the-town-of-joshimath-northern-india-is-sliding/1149>, last accessed on 30.1.2023.
- FRANK, F., MCARDELL, B.W., HUGGEL, C. & VIELI, A., 2015: The importance of erosion for debris flow runout modelling from applications to the Swiss Alps. *Natural Hazards Earth System Sciences Discussions*, **3**, 2379-2417.
- GLIMS CONSORTIUM, 2005: GLIMS Glacier Database. Version 1. Boulder Colorado, USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. <https://doi.org/10.7265/N5V98602>.

- GRUBER, S., 2012: Derivation and analysis of a high-resolution estimate of global permafrost zonation. *The Cryosphere*, **6**, 221-233. <https://doi.org/10.5194/tc-6-221-2012>.
- JONES, N., MANCONI, A. & STROM, A., 2021: Active landslides in the Rogun Catchment, Tajikistan, and their river damming hazard potential. *Landslides*, **18**, 3599-3613.
- LIU, Z., QIU, H., ZHU, Y., LIU, Y., YANG, D., MA, S., ZHANG, J., WANG, Y., WANG, L. & TANG, B., 2022: Efficient Identification and Monitoring of Landslides by Time-Series InSAR Combining Single- and Multi-Look Phases. *Remote Sensing*, **14**(1026).
- MANCONI, A., 2021: How phase aliasing limits systematic space-borne DInSAR monitoring and failure forecast of alpine landslides. *Engineering geology*, **287**(106094).
- PANDEY, V.K., KUMAR, R., SINGH, R., KUMAR, R., RAI, S.C., SINGH, R.P., TRIPATHI, A.K., SONI, V.K., ALI, S.N., TAMANG, D. & LATIEF, S.U., 2021: Catastrophic ice-debris flow in the Rishiganga River, Chamoli, Uttarakhand (India). *Geomatics, Natural Hazards and Risk*, **13**(1), 289-309.
- PASQUALI, P., CANTONE, A., RICCARDI, P., DEFILIPPI, M., OGUSHI, F., GAGLIANO, S. & TAMURA, M., 2014: Mapping of Ground Deformations with Interferometric Stacking Techniques. *Land Applications of Radar Remote Sensing*, InTech.
- SATI, S.P., SHARMA, S., SUNDRIYAL, Y.P., RAWAT, D. & RIYAL, M., 2020: Geo-environmental consequences of obstructing the Bhagirathi River, Uttarakhand Himalaya, India. *Geomatics, Natural Hazards and Risk*, **11**(1), 887-905.
- SCHNEIDER, D. & HAEBERLI, W., 2011. On characteristics and flow dynamics of large rapid mass movements in glacial environments. Doctoral thesis, Department of Geography, University of Zurich.
- SHUGAR, D.H., JACQUEMART, M., SHEAN, D., BHUSHAN, S., UPADHYAY, K., SATTAR, A., SCHWANGHART, W., MCBRIDE, S., VAN WYK DE VRIES, M., MERGILI, M., EMMER, A., DESCHAMPS-BERGER, C., MCDONNELL, M., BHAMBRI, R., ALLEN, S., BERTHIER, E., CARRIVICK, J.L., CLAGUE, J.J., DUKUKIN, M., DUNNING, S.A., FREY, H., GASCOIN, S., HARITASHYA, U.K., HUGGEL, C., KÄÄB, A., KARGEL, J.S., KAVANAUGH, J.L., LACROIX, P., PETLEY, D., RUPPER, S., AZAM, M.F., COOK, S.J., DIMRI, A.P., ERIKSSON, M., FARINOTTI, D., FIDDES, J., GNYAWALI, K.R., HARRISON, S., JHA, M., KOPPES, M., KUMAR, A., LEINSS, S., MAJEED, U., MAL, S., MUHURI, A., NOETZLI, J., PAUL, F., RASHID, I., SAIN, K., STEINER, J., UGALDE, F., WATSON, C.S. & WESTOBY, M.J., 2021: A massive rock and ice avalanche caused the 2021 disaster at Chamoli, Indian Himalaya. *Science*, **373**, 300-306.
- STROZZI, T., KÄÄB, A. & FRAUENFELDER, R., 2004: Detecting and quantifying mountain permafrost creep from in situ inventory, space-borne radar interferometry and airborne digital photogrammetry. *International Journal of Remote Sensing*, **25**(15), 2919-2931.
- VERMA, S., SHARMA, A., YADAVA, P., GUPTA, P., SINGH, J. & PAYRA S., 2022: Rapid flash flood calamity in Chamoli, Uttarakhand region during Feb 2021: an analysis based on satellite data. *Natural Hazards*, **112**, 1379-1393.
- ZHOU, L., WANG, C., TANG, Y., ZHANG, B., ZHANG, H. & DONG, L., 2022: Interferometric SAR Observation of Permafrost Status in the Northern Qinghai-Tibet Plateau by ALOS, ALOS-2 and Sentinel-1 between 2007 and 2021. *Remote Sensing*, **14**(1870).

# Silbermöwe oder Heringsmöwe? - Klassifikation von Seevögeln in UAS-Aufnahmen

KATRIN GÖLZ<sup>1</sup>, SYLVIA HOCHSTUHL<sup>1,2</sup> & ANTJE THIELE<sup>1,2</sup>

*Zusammenfassung: Für eine effiziente Erfassung der Vogelbestände, werden in dieser Arbeit lernbasierte Verfahren für die Klassifikation unterschiedlicher Seevögel in Luftbildaufnahmen getestet und verglichen. Ausgewertet werden zwei Datensätze, welche über Spiekeroog und Baltrum mittels unbemannter Flugsysteme aufgenommen wurden. Für die Unterteilung von Bildausschnitten in die Klassen Seevogel und Hintergrund kommen ein SVM-basierter und ein CNN-basierter Klassifikator zum Einsatz. Beide Ansätze liefern mit Gesamtgenauigkeiten von über 92% sehr gute Ergebnisse, wobei die Klassifikationsleistung des CNNs die des SVM-basierten Verfahrens leicht übersteigt. Die CNN-basierte Klassifikation wird darüber hinaus für die Unterscheidung von bis zu sechs unterschiedlichen Vogelarten eingesetzt. Während Vogelarten, für welche eine ausreichende Anzahl an Trainingsbeispielen verfügbar ist, mit recall-Werten von bis zu 96% erkannt werden, liegen die recall-Werte der im Training unterrepräsentierten Klassen deutlich niedriger, sodass balanced accuracy-Werte von bis zu 62% und 84% erzielt werden.*

## 1 Einleitung

Mehr als 200 Vogelarten sind als Brutvögel in Niedersachsen heimisch. Eine kontinuierliche Erfassung der Daten zu Vorkommen, Verbreitung und Bestandsentwicklung der Brutvögel ist für die Naturschutzarbeit von Behörden und Verbänden essentiell. So dienen die erhobenen Daten als wichtige Entscheidungsgrundlagen bei der Ausweisung von Schutzgebieten, deren Abgrenzung und Zonierung sowie der Planung von Schutzmaßnahmen. Neben der herkömmlichen und aufwändigen Methode, Vögel bei Feldbegehungen zu zählen, werden zunehmend hochaufgelöste Luftbildaufnahmen für die Vogelzählung herangezogen, welche mittels unbemannten Flugsystemen (engl. Unmanned Aircraft System (UAS)) aufgenommen werden. Der Einsatz von UAS-Systemen bietet den Vorteil der flächendeckenden, effizienten und kostengünstigen Datenerfassung, durch welche die Brutvögel laut SARDÀ-PALOMERA et al. (2012) und GRENZDÖRFFER (2013) nicht gestört werden. Für die Vogelzählung werden die Luftbildaufnahmen in der Regel von fachkundigem Personal manuell ausgewertet. Da in großen Brutgebieten teilweise mehrere Tausend Vögel brüten, ist die manuelle Detektion und Klassifikation aller Vogelarten allerdings eine langwierige Aufgabe. Die automatische Bilddatenauswertung hat somit das Potential den Arbeitsaufwand der Vogelzählung deutlich zu reduzieren.

Der erfolgreiche Einsatz automatischer Detektionsverfahren für das Zählen von Vögeln in Luftbildaufnahmen wird in unterschiedlichen Studien demonstriert (GRENZDÖRFFER 2013; DIAZ-DELGADO et al. 2017; AFÁN et al. 2018; HODGSON et al. 2018; RUSH et al. 2018). Dabei werden

---

<sup>1</sup> Karlsruher Institut für Technologie, Institut für Photogrammetrie und Fernerkundung, Englerstr. 7, 76131 Karlsruhe, E-Mail: Katrin.Goelz@gmx.de

<sup>2</sup> Fraunhofer IOSB, Ettlingen, Gutleuthausstr. 1, 76275 Ettlingen, E-Mail: [Sylvia.Hochstuhl, Antje.Thiele]@iosb.fraunhofer.de



bereits teilweise Methoden aus dem Bereich des maschinellen Lernens eingesetzt. So werden in der Arbeit von DIAZ-DELGADO et al. (2017) Möwen in UAS-Bildaufnahmen mittels Support Vector Machine (SVM) und Random Forest detektiert und automatisch gezählt. Die so erzielten Ergebnisse werden anhand von Bodenbeobachtungen validiert und ergeben für beide Ansätze hohe Gesamtgenauigkeiten von 82 % bis 98 %. Damit ist die automatische Klassifikation nur wenig ungenauer als die zeitaufwändige manuelle Auswertung der UAS-Bildaufnahmen. Die Random Forest-basierte Klassifikation wird auch von AFÁN et al. (2018) für die Erkennung des Braunen Sichlers verwendet. In diesem Anwendungsfall ist die Gesamtgenauigkeit mit 46% jedoch unter anderem aufgrund des dunklen Gefieders der Vögel, niedrig. Eine weitere vielversprechende Methode für die automatische Vogelzählung ist die CNN (Convolutional Neural Network)-basierte Klassifikation. So stellen HONG et al. (2019) eine Studie vor, in welcher Vögel in UAS-Aufnahmen mittels der Detektionsnetze Yolo, Faster Region-based CNN, Region-based Fully Convolutional Network, Single Shot MultiBox Detector und Retinanet mit durchschnittlichen Genauigkeiten (average precision) zwischen 85% und 95% erkannt werden. Auch in der Arbeit von KELLENBERGER et al. (2021) bringt die CNN-basierte Detektion von Seevögeln gute Ergebnisse, die durch die Einbeziehung von Vorwissen über die räumliche Verteilung von Vogelkolonien mittels Markov Random Fields noch verbessert werden können.

Für die Detektion von Seevögeln in Brutgebieten auf den ostfriesischen Inseln kommen in dieser Arbeit die SVM-basierte und CNN-basierte Klassifikation zum Einsatz. In den betrachteten Brutgebieten nisten verschiedene Arten von Seevögeln. Um Rückschlüsse auf die Populationsentwicklung der unterschiedlichen Vogelarten zu ziehen, wird in dieser Arbeit neben der reinen Erkennung von Vögeln, die in den angeführten Studien schwerpunktmäßig thematisiert wird, ein weiterer Fokus auf die Klassifikation der Seevogelarten gelegt. Eine spezielle Herausforderung dabei ist die unausgewogene Klassenverteilung. Während für einige Vogelarten mehrere Tausend Trainingsbeispiele zur Verfügung stehen, sind andere Arten nur schwach repräsentiert.

Die vorliegende Arbeit ist folgendermaßen gegliedert: Die Datenerhebung und -aufbereitung der UAS-Aufnahmen wird in Abschnitt 2 vorgestellt. In Abschnitt 3 folgt die Beschreibung der verwendeten Methodik für die automatische Klassifikation. Die erzielten Ergebnisse werden in Abschnitt 4 vorgestellt und diskutiert. Abschließend werden die Kernaussagen in Abschnitt 5 zusammengefasst und ein Ausblick auf zukünftige Arbeiten wird gegeben.

## 2 Datengrundlage

Als Datengrundlage dienen in dieser Arbeit zwei Datensätze des Niedersächsischen Landesbetriebs für Wasserwirtschaft, Küsten- und Naturschutz (NLWKN), welche 2018 und 2021 über zwei verschiedenen Untersuchungsgebieten auf Spiekeroog und Baltrum mittels UAS aufgenommen wurden. Die beflogenen Gebiete sind in Abb. 1 dargestellt. Im Folgenden wird das Vorgehen der Datenerfassung und der visuellen Auswertung der Bilder für die Generierung von Trainingsdaten vorgestellt.

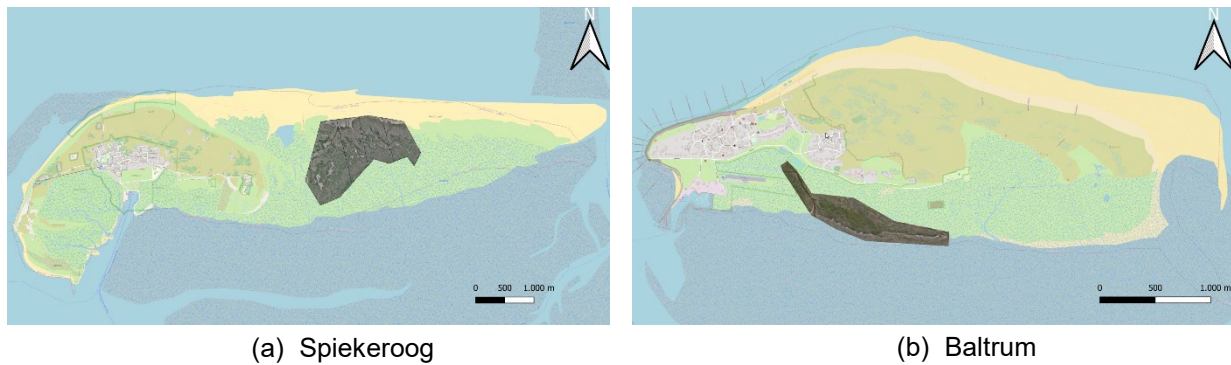


Abb. 1: Aufgenommene Brutgebiete auf Spiekeroog und Baltrum

## 2.1 Datenerfassung

Im Rahmen der UAS-Messkampagnen, deren Ziel insbesondere die Zählung brütender Seevögel ist, wurden die Untersuchungsgebiete jeweils zweimal an einem Tag innerhalb des Brutzeitraums (Mitte April bis Anfang Juli) befliegen. Die zeitlich leicht versetzte Aufnahme des gleichen Gebietes soll dazu dienen brütende von nichtbrütenden Seevögeln zu unterscheiden. Die Befliegungen wurden um die Mittagszeit durchgeführt, um einen möglichst kurzen Schattenwurf der beobachteten Vögel zu gewährleisten. Optimalerweise sollte der Himmel während der Befliegung bewölkt sein, sodass diffuse Lichtverhältnisse vorherrschen. Angaben zu dem verwendeten System für die Bildaufnahme sind Tab. 1 zu entnehmen. Um die Generierung von Orthofotos, welche das beobachtete Gebiet maßstabsgetreu und verzerrungsfrei darstellen, zu ermöglichen, wurden die Gebiete mit einer Bildüberlappung von etwa 80% längs und 60% quer zu der Flugrichtung aufgenommen. Über den GNSS-Empfänger, welcher auf dem UAS angebracht ist, wird die Position zu jedem Aufnahmezeitpunkt ermittelt und mittels Real Time Kinematic (RTK) korrigiert. Damit können die Orthofotos georeferenziert werden.

Abb. 2 zeigt einen Ausschnitt einer UAS-Aufnahme über Baltrum aus dem Jahr 2021. Die auf dem Boden sitzenden Vögel sind in der Regel gut erkennbar. Auffallend ist, dass die Vögel auf hellem, sandigen Boden (im linken Teil der Abbildung) dichter angesiedelt sind als auf den Vegetationsflächen (im rechten Teil der Abbildung). Es handelt sich dabei um die Vogelarten Brandseeschwalbe und Lachmöwe, die verschiedene Bruthabitate präferieren.

Tab. 1: Details zu den Untersuchungsgebieten und genutzten Aufnahmesystemen

	Spiekeroog	Baltrum
Gebietsgröße	1,74 km <sup>2</sup>	0,34 km <sup>2</sup>
Aufnahmezeitpunkt	29.05.2018	28.05.2021
UAS	Delair UX11	DJI Matrice 300 RTK
Flughöhe	75 m	65 m
Bodenauflösung	10,0 mm pro Pixel	7,5 mm pro Pixel



Abb. 2: Orthofotoausschnitt über Baltrum mit Brutkolonien der Brandseeschwalbe und der Lachmöwe

## 2.2 Manuelle Datenauswertung

Die Orthofotos werden mittels manueller, visueller Analyse von Experten des NLWKNs ausgewertet. Dabei wird jeder Seevogel im Orthofoto markiert und seiner entsprechenden Art zugeordnet. Die Vogelarten, die in den erfassten Brutgebieten nisten sind: Der Austernfischer, die Brandseeschwalbe, die Gans, die Heringsmöwe, die Lachmöwe, der Löffler, die rotfüßige Seeschwalbe und die Silbermöwe. Letztendlich liegt es im Ermessen der auswertenden Person, welcher Art ein Individuum zugeordnet wird. Damit ist die zugewiesene Vogelart mit einer Unsicherheit behaftet. Das Vorkommen der unterschiedlichen Vogelarten ist Standort spezifisch. Während im Brutgebiet auf Spiekeroog vor allem Heringsmöwen und Silbermöwen zu finden sind, brüten auf Baltrum vorwiegend Brandseeschwalben und Lachmöwen.

Der manuelle Detektions- und Klassifikationsprozess für das Brutgebiet auf Spiekeroog nimmt ca. 1,5 Wochen in Anspruch. Für das kleinere Brutgebiet auf Baltrum, werden ca. 3 Tage benötigt. Das Ergebnis der manuellen Auswertung dient in dieser Arbeit als Grundlage für die Generierung eines Datensatzes für das Training und die Evaluierung lernbasierter Klassifikatoren, welche Bildausschnitte als *Hintergrund* oder als entsprechenden *Seevogel* einordnen. Dazu werden aus den Orthofotos Bildausschnitte fester Größe extrahiert, die jeweils einen Seevogel abbilden. Der Seevogel liegt dabei im Zentrum des Bildausschnittes und hat eine zufällige Ausrichtung. Zusätzlich werden Bildausschnitte erstellt, welche ausschließlich Hintergrund, sprich keinen Seevogel, abbilden. Um einen möglichst fehlerfreien Trainingsdatensatz zu gewährleisten, werden alle Bildausschnitte und ihre entsprechende Klassenzuordnung händisch kontrolliert. Dabei werden fehlerhaft gekennzeichnete Bildausschnitte oder Bildausschnitte, die Artefakte aufweisen, aussortiert. Die Klassenverteilungen der resultierenden Datensätze sind in Tab. 2 aufgeschlüsselt.

Tab. 2: Art und Anzahl verschiedener Seevögel in den erstellten Datensätzen

Vogelart	Abkürzung	Vorkommen auf Spiekeroog	Vorkommen auf Baltrum
Austernfischer	AF	81	80
Brandseeschwalbe	BS	-	3.261
Gans	G	900	97
Heringsmöwe	HM	9.450	704
Lachmöwe	LM	149	24.300
Löffler	LF	132	60
Rotfüßige Seeschwalbe	RS	-	104
Silbermöwe	SM	2.110	524
Hintergrund	HG	4.336	6.122

### 3 Methodik

Für die Unterteilung von Bildausschnitten in die Klassen *Seevogel* und *Hintergrund* werden zwei lernbasierte Methoden - SVM und CNN - angewendet und verglichen. Während die SVM lernt, Bildausschnitte anhand extrahierter Histogram of Oriented Gradients (HOG)-Deskriptoren zu klassifizieren, lernt das CNN sowohl die Extraktion geeigneter Merkmale als auch die darauf basierende Klassifikation. Im Folgenden werden beide Ansätze näher erläutert.

#### 3.1 Klassifikation mittels SVM

Um charakteristische Bildinformation einzelner Bildausschnitte in kompakter Form darzustellen, kommen in der Regel Merkmalsdeskriptoren zum Einsatz. Zu den gewünschten Eigenschaften eines Deskriptors (auch Merkmal genannt) gehört die Invarianz gegenüber Skalierung, Translation und Rotation, sowie die Robustheit in Hinblick auf Beleuchtungsveränderung und Bildrauschen. In dieser Arbeit wird der sogenannte HOG-Deskriptor verwendet, welcher auf Helligkeitsgradienten innerhalb von Pixelnachbarschaften basiert (DALAL & TRIGGS 2005). Für die Berechnung des Deskriptors werden die RGB-Bilder zunächst in Grauwertbilder umgewandelt und mittels Sobel-Operator werden für jedes Pixel Gradientenbetrag und -orientierung bestimmt. Darauffolgend wird das Bild in mehrere, rechteckige Bereiche gleicher Größe unterteilt, die als Zellen bezeichnet werden. In dieser Arbeit wird dazu eine Zellengröße von  $9 \times 9$  Pixel verwendet. Für jede dieser Zellen wird ein Histogramm berechnet, welches den akkumulierten Gradientenbetrag über die Gradientenorientierung aufträgt. Dies ist in Abb. 3 veranschaulicht. Um den Einfluss von Beleuchtungsänderung innerhalb eines Bildes zu reduzieren, wird eine Normalisierung der Histogrammwerte über einen größeren Bildbereich (Block) durchgeführt. In dieser Arbeit wird dazu eine Blockgröße von  $2 \times 2$  Zellen genutzt. Der HOG-Deskriptor für den gesamten Bildausschnitt ergibt sich aus der Verkettung aller normalisierten Blockhistogramme.

Die HOG-Deskriptoren bilden die Eingangsdaten für die SVM. Innerhalb des Trainingsprozesses wird anhand von Beispielen eine Hyperebene gesucht, welche die HOG-Deskriptoren unterschiedlicher Klassen bestmöglich trennt. Da nicht davon ausgegangen wird, dass die vorliegenden HOG-Deskriptoren unterschiedlicher Klassen linear trennbar sind, wird der sogenannte Kernel-Trick angewendet, wodurch die Merkmale in einen höherdimensionalen Raum transformiert werden, in welchem eine passende Trennebene einfacher gefunden wird. Nach dem Training, wird die SVM genutzt, um ungesehene Bildausschnitte anhand ihrer HOG-Deskriptoren einer Klasse zuzuweisen.

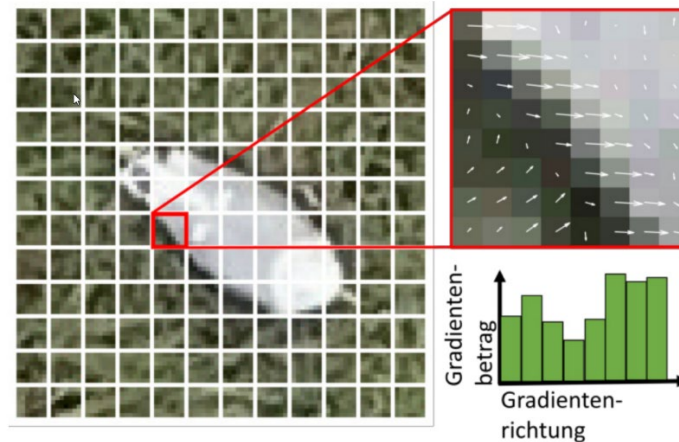


Abb. 3: Erstellung von HOG-Deskriptoren durch Histogrammbildung basierend auf Gradientenbetrag und -orientierung

### 3.2 Klassifikation mittels CNN

Im Gegensatz zu der SVM-basierten Klassifikation entfällt bei der CNN-basierten Klassifikation der Schritt der Merkmalsextraktion, da die RGB-Bilder direkt als Eingangsdaten genutzt werden können. Der Aufbau des CNNs, welches in dieser Arbeit genutzt wird, ist in Abb. 4 dargestellt. Der Bildausschnitt, der als Eingabeschicht dient, wird in der Convolutional-Schicht zunächst mit unterschiedlichen Filterkernen gefaltet. Die Filtergewichte stellen dabei variable Parameter dar, die im Rahmen des Trainingsprozesses optimiert werden. Die resultierenden Merkmale werden normiert (Batch Normalisierung) und es folgt eine nichtlineare Aktivierungsfunktion (ReLU). Durch eine Max-Pooling-Schicht werden die Merkmalsinformationen räumlich aggregiert und anschließend in den vollvernetzten Schichten miteinander verknüpft. Die Softmax-Funktion am Ende des Netzes transformiert die Ausgabe der vorangegangenen Schicht in eine kategoriale Wahrscheinlichkeitsverteilung. Die Optimierung der Netzparameter erfolgt anhand von Trainingsdaten durch Minimierung der Kreuzentropiefunktion. Nach der Trainingsphase können ungesehene Bildausschnitte vorwärts durch das Netz propagiert werden und einer der zuvor definierten Klassen zugewiesen werden.

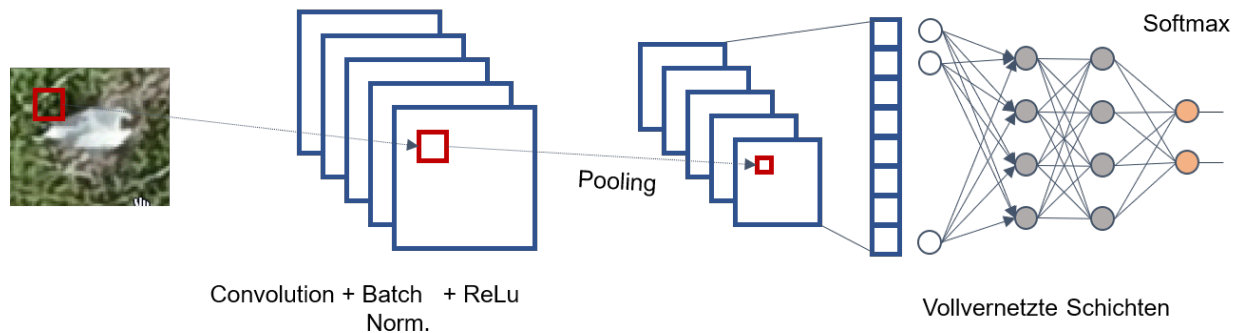


Abb. 4: Architektur des eingesetzten CNNs

### 3.3 Unausgewogene Trainingsdaten

Wie der in Tab. 2 beschriebenen Klassenverteilung des vorliegenden Datensatzes zu entnehmen ist, unterscheidet sich die Anzahl von Trainingsbeispielen pro Vogelart stark. Lernbasierte Modelle, die auf unausgewogenen Daten trainiert werden, neigen dazu überrepräsentierte Klassen aufgrund ihrer höheren Eintrittswahrscheinlichkeit häufiger vorherzusagen, während unterrepräsentierte Klassen teilweise vollständig ignoriert werden. Um dem ungewünschten Effekt einer Klassenüberanpassung entgegen zu wirken, werden im Rahmen der CNN-Klassifikation zwei Gegenmaßnahmen eingesetzt und verglichen. Zum einen werden die Trainingsdaten mittels einer Kombination aus Datenerweiterung und -reduktion ausbalanciert, zum anderen wird eine robuste Fehlerfunktion, der sogenannte Focal Loss, eingesetzt.

#### 3.3.1 Datenerweiterung und -reduktion

Unter Datenerweiterung versteht man die synthetische Vergrößerung von Datensätzen. Dieses Verfahren wird häufig eingesetzt, wenn eine unzureichende Menge an Daten für das Training eines Klassifikators zur Verfügung steht. Aus den vorhandenen Daten werden dabei beispielsweise durch Translation, Rotation und Skalierung synthetische Daten generiert. In dieser Arbeit wird diese Datenerweiterung genutzt, um synthetische Daten für die unterrepräsentierten Klassen zu erstellen. Damit ein Klassengleichgewicht erreicht werden kann ohne einen unverhältnismäßig hohen Anteil synthetischer Daten zu verwenden, wird die Menge der Daten überrepräsentierter Klassen durch eine zufällige Auswahl reduziert.

#### 3.3.2 Focal Loss

Wird für das Training des zuvor vorgestellten CNNs die Kreuzentropie-Fehlerfunktion verwendet, wird der Gesamtfehler durch überrepräsentierte Klassen in einem unausgewogenen Datensatz dominiert. Dadurch werden die trainierbaren Parameter in die Richtung aktualisiert, dass das Modell mehr Vertrauen in die Vorhersage der überrepräsentierten Klassen gewinnt, während die Minderheitsklassen weniger Gewicht bekommen. Die Focal Loss-Funktion, die von LIN et al. (2017) vorgeschlagen wird, stellt eine Modifikation der Kreuzentropie-Fehlerfunktion dar, die dazu führt, dass der Fokus auf schwierige, fehlklassifizierte Beispiele gesetzt wird. Dazu wird ein dynamischer Skalierungsfaktor eingeführt, welcher mit zunehmendem Vertrauen des Modells in die richtige Klasse gegen Null sinkt.

## 4 Ergebnisse

Für das Training und die Evaluation der Klassifikatoren werden die beiden Datensätze (Spiekeroog und Baltrum) in einem Verhältnis von 2:3 in Trainings- und Testdaten aufgeteilt. Das Training und die Evaluation erfolgen für beide Datensätze getrennt. Der Vergleich von SVM- und CNN-basierter Klassifikation wird anhand einer binären Klassifikation ausgewertet, bei welcher ein Bildausschnitt der Klasse *Seevogel* oder *Hintergrund* zugewiesen wird. Die Auswertung der Multiklassen-Klassifikation, bei welcher zusätzlich die Vogelart bestimmt wird, beschränkt sich auf die CNN-basierte Klassifikation. Für die Darstellung und Bewertung der Ergebnisse, die auf den Testdaten erzielt werden, wird jeweils die Konfusionsmatrix herangezogen sowie die daraus abgeleiteten Kenngrößen recall, precision, Gesamtgenauigkeit und balanced accuracy.

### 4.1 Binäre Klassifikation

Die Ergebnisse der binären Klassifikation mittels SVM und CNN sind in Tab. 3 (Spiekeroog) und Tab. 4 (Baltrum) dargestellt. In der letzten Zeile sind jeweils die recall-Werte, in der letzten Spalte die precision-Werte und in der rechten unteren Ecke die Gesamtgenauigkeit aufgetragen. Beide Methoden erzielen mit über 95% sehr hohe recall-Werte für die Klasse *Seevogel* sowie sehr hohe precision-Werte (>85%), wobei der CNN-Klassifikator um wenige Prozentpunkte besser abschneidet. Größere Unterschiede zwischen den beiden Methoden zeigen sich in der Erkennung der Klasse *Hintergrund*. Während die SVM 81,1% bzw. 88,9% der *Hintergrund*-Bilder als solche erkennt, erzielt das CNN recall-Werte von 91,4% bzw. 96,1%. Die Ergebnisse zeigen, dass die CNN-basierte Methode der SVM überlegen ist. Die durch das CNN automatisch gelernten Merkmale scheinen für die Unterscheidung von *Seevogel* und *Hintergrund* somit besser geeignet zu sein als die HOG-Merkmale.

Tab. 3: Testergebnisse, die auf dem Spiekeroog-Datensatz mittels SVM- und CNN-basierter Klassifikation erzielt werden

		Referenz		
		<i>Hintergrund</i>	<i>Seevogel</i>	
Prädiktion	<i>Hintergrund</i>	2.138	366	0,85
	<i>Seevogel</i>	498	7.982	0,94
		0,81	0,95	0,92

SVM-basierte Klassifikation

		Referenz		
		<i>Hintergrund</i>	<i>Seevogel</i>	
Prädiktion	<i>Hintergrund</i>	2.410	158	0,94
	<i>Seevogel</i>	226	8.190	0,97
		0,91	0,98	0,96

CNN-basierte Klassifikation

Tab. 4: Testergebnisse, die auf dem Baltrum-Datensatz mittels SVM- und CNN-basierter Klassifikation erzielt werden

		Referenz							
		<i>Hintergrund</i>	<i>Seevogel</i>						
Prädiktion	<i>Hintergrund</i>	3.266	415	0,89	Prädiktion	<i>Hintergrund</i>	3.529	94	0,97
	<i>Seevogel</i>	407	16.640	0,98		<i>Seevogel</i>	144	16.961	0,99
		0,89	0,98	0,96			0,96	0,98	0,99
SVM-basierte Klassifikation					CNN-basierte Klassifikation				

## 4.2 Multiklassen-Klassifikation

Bei der Multiklassen Klassifikation wird zusätzlich zwischen verschiedenen Vogelarten unterschieden. Für den Spiekeroog-Datensatz erfolgt eine Einteilung in die sieben Klassen *Hintergrund* (HG), *Heringsmöwe* (HM), *Silbermöwe* (SM), *Austernfischer* (AF), *Gans* (G), *Löffler* (LF) und *Lachmöwe* (LM). Die im Baltrum-Datensatz abgebildeten Seevögel werden in die Klassen *Brandseeschwalbe* (BS), *Lachmöwe* (LM) und *andere Arten* (AA) unterteilt. Die Zusammenfassung unterschiedlicher Arten in die Klasse AA erfolgt aufgrund der niedrigen Anzahl verfügbarer Beispielfelder pro Klasse.

Für die CNN-basierte Multiklassen-Klassifikation kann ein ein- oder mehrstufiges Verfahren genutzt werden. Im einstufigen Verfahren, wird ein Modell trainiert, welches direkt zwischen sieben bzw. vier Klassen unterscheidet, welches die verschiedenen Vogelarten sowie die Klasse Hintergrundklasse HG beinhalten. Bei der mehrstufigen Klassifikation werden hingegen zwei Modelle trainiert. Das erste Modell wird für die binäre Klassifikation (*Seevogel*, *Hintergrund*) eingesetzt. Alle Bildausschnitte, die durch das erste Modell als *Seevogel* klassifiziert werden, werden von einem zweiten Modell der abgebildeten Vogelart zugeordnet. Durchgeführte Versuchsreihen haben gezeigt, dass sich die Klassifikationsleistung des einstufigen Verfahrens im Vergleich zu dem zweistufigen Verfahren nur geringfügig verschlechtert oder in manchen Fällen sogar verbessert. Da der Trainings- und Testaufwand des einstufigen Verfahrens deutlich geringer ist, wird dieses in den folgenden Analysen verwendet.

Die CNN-basierte Klassifikation der Vogelart wird in den folgenden drei verschiedenen Konfigurationen durchgeführt:

1. Training des CNNs mit Kreuzentropie-Fehlerfunktion anhand einer stratifizierten Stichprobe des jeweiligen Datensatzes.
2. Training des CNNs mit Kreuzentropie-Fehlerfunktion anhand ausbalancierter Trainingsdaten.
3. Training des CNNs mit Focal Loss-Funktion anhand einer stratifizierten Stichprobe des jeweiligen Datensatzes.



Tab. 5: Konfusionsmatrizen der CNN-basierten Klassifikation des Spiekeroog-Datensatzes

		Referenz								
		<i>HG</i>	<i>HM</i>	<i>SM</i>	<i>AF</i>	<i>G</i>	<i>LF</i>	<i>LM</i>		
Prädiktion	<i>HG</i>	2.493	102	16	6	86	4	2	0,92	Stratifizierte Trainingsdaten <b>Balanced accuracy: 0,60</b>
	<i>HM</i>	34	5.183	294	10	228	0	65	0,89	
	<i>SM</i>	10	272	931	0	16	9	17	0,74	
	<i>AF</i>	5	6	0	20	53	0	0	0,24	
	<i>G</i>	21	62	1	13	155	0	1	0,61	
	<i>LF</i>	39	0	4	0	0	66	0	0,61	
	<i>LM</i>	0	45	19	0	2	0	5	0,07	
		0,96	0,91	0,74	0,42	0,29	0,83	0,06	0,83	
Prädiktion	<i>HG</i>	2.508	261	38	5	112	8	2	0,85	Ausbalancierte Trainingsdaten <b>Balanced accuracy: 0,62</b>
	<i>HM</i>	26	4.105	197	2	140	0	40	0,91	
	<i>SM</i>	10	703	945	0	24	6	21	0,55	
	<i>AF</i>	0	23	0	23	68	0	0	0,20	
	<i>G</i>	29	215	4	19	192	1	1	0,42	
	<i>LF</i>	26	6	6	0	0	64	0	0,63	
	<i>LM</i>	3	357	76	0	4	0	25	0,05	
		0,96	0,72	0,75	0,46	0,36	0,81	0,28	0,75	
Prädiktion	<i>HG</i>	2.426	51	16	2	47	9	1	0,95	Stratifizierte Trainingsdaten & Focal Loss-Funktion <b>Balanced accuracy: 0,60</b>
	<i>HM</i>	76	5.202	373	9	242	1	65	0,87	
	<i>SM</i>	23	286	829	0	21	5	12	0,70	
	<i>AF</i>	5	0	0	23	49	0	0	0,30	
	<i>G</i>	23	74	1	15	179	0	0	0,61	
	<i>LF</i>	44	0	4	0	0	64	0	0,57	
	<i>LM</i>	5	57	43	0	2	0	11	0,09	
		0,93	0,92	0,65	0,46	0,33	0,81	0,13	0,82	

Tab. 6: Konfusionsmatrizen der CNN-basierten Klassifikation des Baltrum-Datensatzes

		Referenz					Referenz					
		<i>HG</i>	<i>BS</i>	<i>LM</i>	<i>AA</i>		<i>HG</i>	<i>BS</i>	<i>LM</i>	<i>AA</i>		
Prädiktion	<i>HG</i>	3515	16	58	72	0,96	<i>HG</i>	3541	16	102	138	0,93
	<i>BS</i>	59	1691	235	62	0,83	<i>BS</i>	77	1767	656	87	0,68
	<i>LM</i>	18	225	14068	317	0,96	<i>LM</i>	7	151	13210	154	0,98
	<i>AA</i>	81	25	219	489	0,60	<i>AA</i>	48	23	612	561	0,45
		0,96	0,86	0,96	0,52	0,93		0,96	0,90	0,91	0,60	0,91

Stratifizierte Trainingsdaten                      Ausbalancierte Trainingsdaten

**Balanced accuracy: 0,83**    **Balanced accuracy: 0,84**

		Referenz				
		<i>HG</i>	<i>BS</i>	<i>LM</i>	<i>AA</i>	
Prädiktion	<i>HG</i>	3442	57	87	82	0,94
	<i>BS</i>	121	1680	365	80	0,75
	<i>LM</i>	55	204	13953	400	0,95
	<i>AA</i>	55	16	175	378	0,61
		0,94	0,86	0,96	0,40	0,92

Stratifizierte Trainingsdaten & Focal Loss-Funktion

**Balanced accuracy: 0,78**

Die auf den Testdaten erzielten Ergebnisse der drei Ansätze sind in Form von Konfusionsmatrizen in Tab. 5 für den Spiekeroog-Datensatz und in Tab. 6 für den Baltrum-Datensatz dargestellt. Zusätzlich ist jeweils der balanced accuracy-Wert angegeben, welcher sich aus der Mittelung der klassenspezifischen recall-Werte ergibt und für die Bewertung der Ergebnisse nicht ausbalancierter Testdaten geeignet ist. Für den Spiekeroog-Datensatz zeigt sich, dass der Hintergrund (*HG*) sowie die Klassen *HM* und *SM* mit recall-Werten von 96%, 91% und 74% sowie mit hohen precision-Werten (>75%) zuverlässig klassifiziert werden. Wie Tab. 2 zu entnehmen ist, sind diese Klassen in den zugrunde liegenden Trainingsdaten gut repräsentiert. Der Löffler (*LF*), der durch seine markante helle Federung eine homogene und gut abgegrenzte Klasse darstellt, wird trotz weniger Trainingsbeispiele gut erkannt. Die Klassen *AF*, *G* und insbesondere *LM*, die in den Trainingsdaten unterrepräsentiert sind, werden hingegen nur selten richtig erkannt und häufig der überrepräsentierten Klasse *HM* zugeordnet. Insgesamt ergibt sich eine Gesamtgenauigkeit von 83% und ein balanced accuracy-Wert von 60%. Wird das Training mit ausbalancierten Daten durchge-

führt, werden die zuvor unterrepräsentierten Klassen (*AF*, *G* und *LM*) insgesamt häufiger prädiert. Die recall-Werte verbessern sich um 4%, 7% und 22%, liegen allerdings noch immer deutlich unter 50%. Gleichzeitig verschlechtert sich der recall-Wert der Klasse *HM*, für welche die Anzahl der Trainingsbeispiele reduziert wurde von 92% auf 72%. Insgesamt ergibt sich im Vergleich zu dem Einsatz stratifizierter Trainingsdaten ein leicht verbesserter balanced accuracy-Wert von 62%. Für die dritte Trainingskonfiguration (Focal Loss-Funktion und stratifizierte Stichprobe), ist die Verbesserung der recall-Werte unterrepräsentierter Klassen schwächer ausgeprägt. Der balanced accuracy-Wert wird nicht verbessert und liegt bei 60%.

Die Auswertung der Klassifikationsergebnisse für den Baltrum-Datensatz (Tab. 6) bestätigt, dass die Nutzung ausbalancierter Daten zu dem höchsten balanced accuracy-Wert (84%) führt, während das Training mit der Focal Loss-Funktion (78%) gegenüber der Kreuzentropie-Fehlerfunktion (83%) insgesamt keinen Vorteil bringt. Für alle drei Trainingskonfigurationen zeigt sich, dass die beiden Hauptklassen *BS* und *LM* sehr gut voneinander getrennt werden können, wobei dies mit dem Training auf einer stratifizierten Stichprobe am besten gelingt. Problematisch ist jeweils die Klassifikation der Klasse *AA*. Als Grund hierfür lässt sich neben der vergleichsweise geringen Anzahl von Trainingsbeispielen, die hohe Heterogenität dieser Klasse, durch die Zusammenfassung unterschiedlicher Vogelarten, anführen.

## 5 Fazit und Ausblick

In dieser Arbeit wurden überwachte Lernverfahren eingesetzt und verglichen, welche das Potential bieten die Erkennung und Klassifikation von Seevögeln in UAS-Aufnahmen zu automatisieren. Für eine binäre Einteilung von Bildausschnitten in die Klassen *Seevogel* oder *Hintergrund* wurde eine SVM-basierte Klassifikation basierend auf HOG-Merkmalen und ein flaches CNN angewendet. Beide Ansätze liefern sehr gute Ergebnisse, wobei die Klassifikationsleistung des CNNs, mit Gesamtgenauigkeiten von bis zu 99%, der SVM, mit einer Gesamtgenauigkeit von bis zu 96%, leicht überlegen ist. Das CNN wurde des Weiteren für die Klassifikation von bis zu sechs verschiedenen Vogelarten eingesetzt. Dabei hat sich gezeigt, dass Vogelarten, für die eine gute Datengrundlage im Sinne umfangreicher Trainingsdaten verfügbar ist, zuverlässig erkannt und unterschieden werden können. In den Trainingsdaten unterrepräsentierte Vogelarten, die zu dem eine hohe Varianz innerhalb der Klasse aufweisen, werden von dem CNN-Modell seltener erkannt und in der Regel den stark repräsentierten Klassen zugewiesen. Um die recall-Werte unterrepräsentierter Klassen zu erhöhen, wurde das CNN zum einen mit einem ausbalancierten Datensatz und zum anderen unter Verwendung der Focal Loss-Funktion trainiert. Die Ergebnisse haben gezeigt, dass die Focal Loss-Funktion die bestehende Herausforderung nicht löst, während sich für das Training mit ausbalancierten Daten eine leichte Verbesserung der balanced accuracy-Werte ergibt. Anstelle der Datenerweiterung und -reduktion die in dieser Arbeit betrachtet wurde, empfiehlt sich für zukünftige Arbeiten unterrepräsentierte Klassen durch Hinzunahme weiterer UAS-Daten (z.B. aus Messkampagnen vergangener Jahre) zu stärken. Auf diese Weise ergeben sich mehr echte Trainingsdaten, welche die CNN-basierte Klassifikation robuster werden lassen. Ein weiterer Fokus zukünftiger Arbeiten besteht darüber hinaus in der Anwendung des Klassifikators auf das gesamte

Befliegungsgebiet. Durch den Abgleich von Klassifikationsergebnissen auf zwei zeitlichen versetzten Aufnahmen, lassen sich Brutvögel identifizieren und automatisch zählen.

## 6 Danksagung

Für die Beratung sowie die Bereitstellung und die manuelle Auswertung der in dieser Arbeit genutzten UAS-Aufnahmen möchten wir uns herzlich bei Holger Dirks und Birgit Hepner des NLWKN bedanken.

## 7 Literaturverzeichnis

- AFÁN, I., MÁÑEZ, M. & DIAZ-DELGADO, R., 2018: Drone Monitoring of Breeding Waterbird Populations: The Case of the Glossy Ibis. *Drones*, **2**(4), 42-55.
- DALAL, N. & TRIGGS, B., 2005: Histograms of Oriented Gradients for Human Detection. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, **1**, 886-893.
- DIAZ-DELGADO, R., MÁÑEZ, M., MARTÍNEZ, A., CANAL, D., FERRER, M. & ARAGONÉS, D., 2017: Using UAVs to Map Aquatic Bird Colonies. *The Roles of Remote Sensing in Nature Conservation*, 277-291.
- GRENDÖRFFER, G., 2013: UAS-based Automatic Bird Count of a Common Gull Colony. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, **40**(1/W2), 169-174.
- HODGSON, J., MOTT, R., BAYLIS, S., PHAM, T., WOTHERSPOON, S., KILPATRICK, A., RAJA SEGARAN, R., REID, I., TERAUDS, A. & KOH, L., 2018: Drones Count Wildlife More Accurately and Precisely than Humans. *Methods in Ecology and Evolution*, **9**(5), 1160-1167.
- HONG, S. J., HAN, Y., KIM, S. Y., LEE, A. Y. & KIM, G., 2019: Application of Deep-learning Methods to Bird Detection Using Unmanned Aerial Vehicle Imagery. *Sensors*, **19**(7), 1651-1667.
- KELLENBERGER, B., VEEN, T., FOLMER, E., & TUIA, D., 2021: 21 000 Birds in 4.5 h: Efficient Large-scale Seabird Detection with Machine Learning. *Remote Sensing in Ecology and Conservation*, **7**(3), 445-460.
- LIN, T., GOYAL, P., GIRSHICK, R., HE, K. & DOLLAR, P., 2017: Focal Loss for Dense Object Detection. *Proceedings of the IEEE International Conference on Computer Vision*, 2980-2988.
- RUSH, G. P., CLARKE, L. E., STONE, M. & WOOD, M. J., 2018: Can Drones Count Gulls? Minimal Disturbance and Semiautomated Image Processing with an Unmanned Aerial Vehicle for Colony-Nesting Seabirds. *Ecology and Evolution*, **8**(24), 12322-12334.
- SARDÀ-PALOMERA, F., BOTA, G., VINOLO, C., PALLARES O. SAZATORNIL, V., BROTONS, L., GOMARIZ, S. & SARDA, F., 2012: Fine-Scale Bird Monitoring from Light Unmanned Aircraft Systems. *Ibis*, **154**(1), 177-183.

# Ko-Registrierung von Sentinel-1 und Sentinel-2 Zeitreihen für die Detektion von Trinkwasserleckagen: ein Methodenvergleich

LINA E. BUDDE<sup>1</sup>, DOROTA IWASZCZUK<sup>1</sup>, LAYTH SAHIB<sup>2</sup> & JÜRGEN BORN<sup>2</sup>

*Zusammenfassung: Leckagen in Fernwasserleitungen in ländlichen Gebieten verursachen große Verluste an Trinkwasser. Ein kontinuierliches Monitoring und damit frühzeitige Erkennung solcher Leckagen ist ohne den Einsatz von Fernerkundung sehr zeit- und kostenintensiv. Durch die gemeinsame Auswertung von Zeitreihen aus den optischen Sentinel-2 Daten als auch SAR Daten von Sentinel-1 können solche Leckagen lokalisiert werden. Eine wichtige Voraussetzung dafür ist genaue Ko-Registrierung der Daten. Daher wurden bestehende Verfahren qualitativ ausgewertet. Sowohl mit der Google Earth Engine als auch mit dem GeFolki Tool kann eine subpixel genaue Ko-Registrierung durchgeführt werden. Dabei zeichnet sich das GeFolki Tool im Vergleich zur Google Earth Engine durch seine Transparenz aus.*

## 1 Einleitung

Der Bedarf an Wasser insbesondere in der Industrie steigt bei gleichzeitig hohen Wasserverlusten beispielsweise durch Verschmutzung (SCHUSTER 2022). Durch den fortschreitenden Klimawandel sind zudem bereits auch in Deutschland Regionen von Dürre bedroht und verringern dadurch zusätzlich die Trinkwasservorkommen (BOEING et al. 2022; KÄSTNER 2023). Umso wichtiger ist es daher, Wasser zu sparen. Während Wasserrohrbrüche in städtischen Gebieten schnell durch überflutete Straßen auffallen, entstehen bei Leckagen in ländlichen Gebieten unbemerkt große Trinkwasserverluste (LAH et al. 2018). Insbesondere bodengestützte Messmethoden wie beispielsweise Druckmessungen bedeuten einen hohen finanziellen Aufwand, ohne, dass dabei zuverlässig kleinere Lecks erkannt und vor allem lokalisiert werden können (LAH et al. 2018). Die Fernerkundung kann jedoch zur Erkennung und Lokalisierung wesentlich beitragen. Während luftgestützte Ansätze durch ihre hohe Auflösung einen Vorteil bieten (KRAPEZ et al. 2022), ermöglichen die frei verfügbaren Sentinel-1 und Sentinel-2 Daten ein kontinuierliches, kostengünstiges und flächendeckendes Monitoring der Wasserleitungen. Im Rahmen eines Projektes der LOEWE Förderlinie 3 wurde daher in Zusammenarbeit mit dem Unternehmen Spatial Business Integration GmbH ein System zur Lokalisierung solcher Leckagen entwickelt. Ein zentraler Bestandteil bildet dafür die Auswertung von Zeitreihen. Diese setzt eine hohe Positionsgenauigkeit der Satellitendaten voraus. Um diese zu erreichen werden die Daten zueinander ko-registriert. Bekannte Verfahren für die Ko-Registrierung beruhen auf Intensitäten oder identischen Punkten (TONDEWAD & DALE 2020). Die Anzahl dauerhafter markanter Punkte in ländlichen Gebieten, die bei den Leckagen im Fokus

---

<sup>1</sup> Technische Universität Darmstadt, Fachbereich Bau- und Umweltingenieurwissenschaften, Fachgebiet Fernerkundung und Bildanalyse, Franziska-Braun-Str. 7, D-64287 Darmstadt, E-Mail: [lina.budde, dorota.iwaszczuk]@tu-darmstadt.de

<sup>2</sup> Spatial Business Integration GmbH, Marienburgstr. 27, D-64297 Darmstadt, E-Mail: [l.y.sahib, juergen.born]@spatial-business-integration.com

stehen, ist jedoch gering. Intensitätsbasierte Ansätze hingegen hängen stark von der jeweiligen Beleuchtung ab. Da somit jedes Verfahren Nachteile besitzt, ist es wichtig Qualitätsangaben zu haben. Bestehende Tools für die Ko-Registrierung liefern allerdings keine einheitlichen Ausgaben und nur begrenzt qualitative Informationen. Insbesondere bei fehlenden Referenzinformationen wie Festpunkten, stellt die qualitative und vergleichbare Auswertung eine Herausforderung dar (USS et al. 2016). Daher wird untersucht, inwieweit bekannte Tools für die Ko-Registrierung im ländlichen Raum im Subpixelbereich geeignet sind und qualitativ ausgewertet werden können.

## 2 Methodik

Für Zeitreihenauswertungen werden zwei verschiedene Ko-Registrierungstools erprobt: die Google Earth Engine (GEE)<sup>3</sup> und GeFolki<sup>4</sup>. Darüber hinaus wird ein Deep Learning Ansatz hinsichtlich seiner Anwendbarkeit geprüft. Ein wichtiger Bestandteil der Untersuchungen ist die Gewinnung von Qualitätsinformation aus den verwendeten Verfahren.

### 2.1 Toolboxen

Die GEE basiert auf der Korrelation der einzelnen gewählten Bildpaare. Dadurch wird eine große Ähnlichkeit der Bilder vorausgesetzt (GOOGLE EARTH ENGINE 2021). Zudem wird in der Beschreibung der GEE angegeben, dass nur ein kleiner Bildversatz bestimmt werden kann. Zusätzlich wird ein Rubber-Sheet Verfahren eingesetzt, welches lokale Bildverzerrungen korrigiert. Im Gegensatz dazu basiert GeFolki auf der Bestimmung eines optischen Flusses (BRIGOT et al. 2016). Die Grundannahme liegt in diesem Fall bei der Invarianz der Pixelintensitäten zwischen zwei Aufnahmen. Aus einer Veränderung der Pixelintensität ergibt sich die gesuchte geometrische Verschiebung. Um diese Verschiebungen auch bei heterogenen Fernerkundungsdaten einsetzen zu können, werden zusätzliche Prozessierungsschritte wie eine Textur- und Kontrastanpassung vorgenommen (BRIGOT et al. 2016). Darüber hinaus wird ein fensterbasierter und somit lokaler Ansatz gewählt. Verwendet werden jeweils die Standardeinstellungen.

### 2.2 Deep Learning

Für die Ko-Registrierung von optischen und SAR Daten mittels Deep Learning stehen verschiedene Ansätze zur Verfügung. Im Gegensatz zu den intensitätsbasierten Verfahren von GEE und GeFolki, beruhen vielversprechende Ansätze im Bereich Deep Learning auf der Suche nach identischen Punkten in SAR-optischen Bildpaaren. Zu dieser Gruppe gehört auch das in diesem Projekt getestete Verfahren SOMatch<sup>5</sup> von HUGHES et al. (2020), bei dem mittels Deep Learning nach korrespondierenden Punkten gesucht wird. Zwei vom Aufbau identische Netze extrahieren jeweils getrennt voneinander in einem optischen beziehungsweise SAR Bild Regionen, die für die Punktzuordnung geeignet sind. Ein drittes Netz dient der genauen

---

<sup>3</sup> <https://earthengine.google.com/>

<sup>4</sup> <https://github.com/aplyer/gefolki>

<sup>5</sup> <https://github.com/system123/SOMatch>

Korrespondenzsuche basierend auf den zuvor gewählten Regionen. Dabei werden jeweils kleinere optische Bildausschnitte mit größeren SAR Ausschnitten korreliert. Im Anschluss erfolgt mit einem vierten Netz eine Ausreißersuche. Da keine trainierten Modelle für die verwendeten Netzwerke vorliegen, müssen diese mit bereits registrierten SAR-optischen Bildpaaren trainiert werden. Nach dem Training können entsprechend der Abbildung 1 korrespondierende Punkte bestimmt werden.

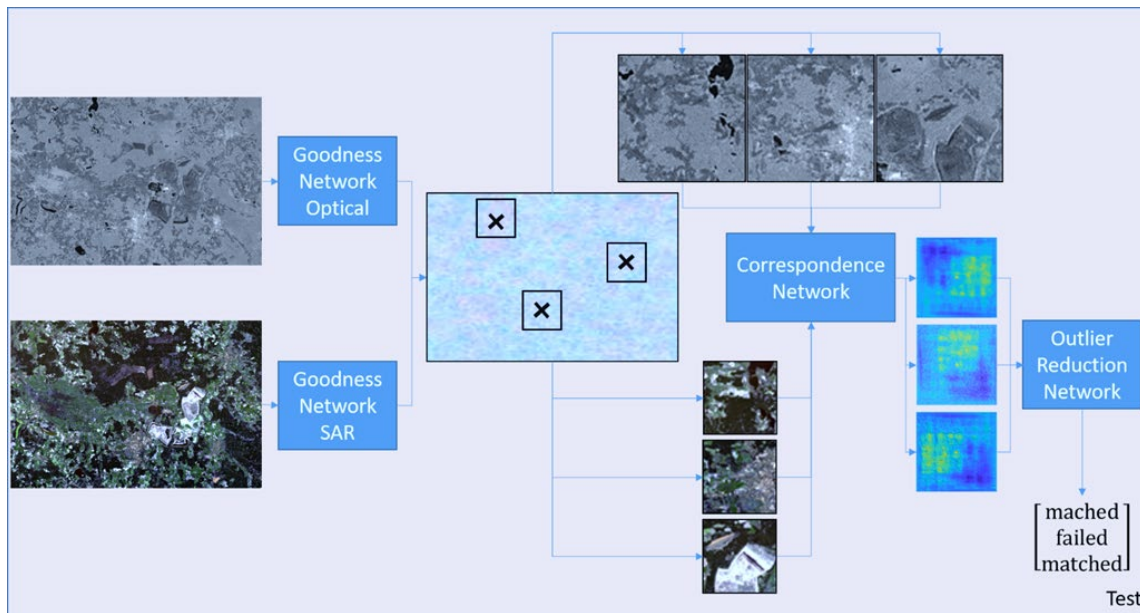


Abb. 1: Schematischer Ablauf der Testphase, bei der ausgehend von einer größeren Bildszene geeignete Bildausschnitte für die Korrespondenzsuche gefunden werden (Goodness Network). Im Anschluss wird die Korrelation zwischen den SAR und optischen Bildausschnitten bestimmt (Correspondence Network) und auf Ausreißer geprüft (Outlier Reduction Network). Adaptiert von HUGHES et al. (2020).

Da sich dieser Deep Learning Prozess auf die Suche nach Korrespondenzen beschränkt, ist die eigentliche Ko-Registrierung nicht enthalten. Infolgedessen erfolgt auf Basis der gefundenen Punktkoordinaten beispielsweise mittels Affintransformation und Ausgleichsrechnung die Bestimmung der Bildtransformationen.

### 2.3 Qualitätsbestimmung

Für eine absolute Genauigkeitsbestimmung der Georeferenzierung fehlen in der Regel gut signalisierte Festpunkte. Daher werden Qualitätsinformation über die Ko-Registrierung intrinsisch aus den Verfahren selbst gewonnen. Somit werden die Abweichungen innerhalb des Ko-Registrierungsvorgangs analysiert. In diesem Fall wird die Annahme verwendet, dass nach einer erfolgreichen Ko-Registrierung keine weitere Transformation notwendig ist. Sowohl bei GEE als auch bei GeFolki wird die Transformation in Form einer pixelweisen Translation  $x, y$  ausgegeben. Diese Translationsvektoren geben die Verschiebung für die Ko-Registrierung an. Treten jedoch Translationen zwischen Referenzbild und bereits ko-registriertem Bild auf, kann aus diesen

verbliebenen Abweichungen  $dx$ ,  $dy$  eine Fehlerabschätzung durchgeführt werden. Die RMSE Werte in x-Richtung werden wie folgt aus den pixelweisen Translationen  $dx_i$  bestimmt:

$$RMSE_{dx} = \sqrt{\sum_i (dx_i)^2} ;$$

die Berechnung in y-Richtung erfolgt äquivalent. Darüber hinaus kann zusätzlich die bei der GEE mitgelieferte Korrelationskonfidenz zur Bewertung herangezogen werden. Für den Vergleich der Verfahren fließen in die Bewertung zusätzlich die Geschwindigkeit, Nutzerfreundlichkeit und Transparenz der Tools ein.

### 3 Versuchsaufbau

#### 3.1 Datensätze

Für den Vergleich der Ko-Registrierungstools wird ein eigener Datensatz bestehend aus einer Sentinel-1 und Sentinel-2 Zeitreihe zwischen Juli und Anfang September 2020 im Gebiet nordwestlich von Darmstadt mit einer Bildausschnittgröße von  $1274 \times 588$  Pixel verwendet. Eine Übersicht über die verschiedenen Zeitreihen ist der Tabelle 1 zu entnehmen. Durch Wolkenbedeckung stehen insgesamt fünf geeignete Aufnahmezeitpunkte bei den optischen Daten zur Verfügung. Für den gleichen Zeitraum wurden insgesamt elf SAR Aufnahmen mit der relativen Orbitnummer 66 eingesetzt. Für die Ko-Registrierung werden für Sentinel-2 nur die RGB Bänder verwendet. Die Sentinel-1 Daten werden entsprechend zum SEN12 Benchmark Datensatz (SCHMITT et al. 2018) vorverarbeitet. Sowohl bei Sentinel-1 als auch bei Sentinel-2 wird eine Bodenauflösung von 10 m verwendet. Zusätzlich wird der SEN12 Benchmark Datensatz (SCHMITT et al. 2018) für das Training des Deep Learning Verfahrens eingesetzt.

Tab. 1: Sentinel-2 (S2) und Sentinel-1 (S1) Zeitreihen für das Untersuchungsgebiet Darmstadt Nordwest. Angegeben sind jeweils Tag und Monat der Aufnahme. Alle Aufnahmen im Untersuchungsgebiet stammen aus dem Jahr 2020.

Bild Nr.	1	2	3	4	5	6	7	8	9	10	11
S2	03.07.	13.07.	23.07.	07.08.	12.08.						
S1	04.07.	10.07.	16.07.	22.07.	28.07.	03.08.	09.08.	15.08.	21.08.	27.08.	02.09.

#### 3.2 Ko-Registrierungsstrategien

Für die Untersuchungen werden verschiedene Ko-Registrierungsstrategien angewendet. Für die Ko-Registrierung von Zeitreihen wird in der ersten Variante „Pfad A“ für jeden Zeitpunkt  $t_i$  das bereits ko-registrierte Bild vom Zeitpunkt  $t_{i-1}$  als Referenzbild verwendet. In der zweiten Variante „Pfad B“ werden alle Zeitpunkte auf dieselbe Referenz  $t_0$  transformiert. Bei der multi-modalen Ko-Registrierung werden SAR-optische Bildpaare ohne zeitliche Komponente ko-registriert.



## 4 Ergebnisse

Nachfolgend sind die Ergebnisse der Versuchsreihen aufgeführt. Bei den Zeitreihenuntersuchungen werden die in 3.2 beschriebenen Strategien jeweils mit GeFolki und GEE ausgeführt und diese qualitativ nach 2.3 ausgewertet.

### 4.1 Vergleich der Ko-Registrierungstools

Zusätzlich zu qualitativen Auswertung werden die eingesetzten Tools hinsichtlich ihrer Bedienbarkeit und ihre Eignung für den praktischen Einsatz im Zusammenhang der Leckage Detektion bewertet. Dafür werden folgende Kriterien berücksichtigt: Geschwindigkeit, Nutzerfreundlichkeit und Transparenz. Tabelle 2 zeigt eine Übersicht über die Kriterien und die verwendeten Tools. Im Gegensatz zu GeFolki und SOMatch ist bei GEE weder die genaue Vorgehensweise beschrieben noch die Implementierung einsehbar. GeFolki weist durch eine Integration sowohl in Python als auch Matlab eine hohe Flexibilität auf. Zudem besitzt GeFolki eine hohe Prozessierungsgeschwindigkeit, welche insbesondere bei großen Datenmengen vorteilhaft ist. Werden die Ko-Registrierungsergebnisse direkt in der GEE weiterverwendet, wird ebenfalls nur eine kurze Verarbeitungszeit benötigt. Jedoch wird ein Wechsel zu einem anderen Verarbeitungstool gewünscht, verlangsamt ein dafür erforderlicher Export der Ko-Registrierung den Verarbeitungsprozess. Bei SOMatch liegt nur eine datensatzspezifische Implementierung vor, die entsprechend für die eigenen Eingabedaten angepasst werden muss. Dadurch werden Programmierkenntnisse in Python und Erfahrung mit Deep Learning Verfahren vorausgesetzt. Zudem ist keine Ende-zu-Ende Implementierung für den Gesamtprozess vorhanden.

Tab. 2: Übersicht über die verwendeten Ko-Registrierungstools und deren Bewertung.

	GEE	GeFolki	SOMatch
Geschwindigkeit	Es werden keine eigenen Rechenkapazitäten benötigt, Export in Google Drive eher langsam	Schnell, abhängig von verwendetem System	Hoher zeitlicher Aufwand für das Training
Nutzerfreundlichkeit	Benötigt Google Account, mit wenigen Befehlen umsetzbar	Matlab auch mit wenig Programmierkenntnis verwendbar, Python Implementierung benötigt mehr Kenntnisse	Viel Programmieraufwand, nur Korrespondenzsuche implementiert
Transparenz	Black Box	Open Source Code, Paper	Open Source Code, Paper

### 4.2 Sentinel-2 Zeitreihe

Bei der Auswertung der Ko-Registrierung der Sentinel-2 Zeitreihe wird durchschnittlich ein Versatz von circa 6 m bei GeFolki und 7 m bei GEE zwischen Referenz und unregistriertem Bild ermittelt. Abbildung 2 zeigt die Fehlerabschätzung nach der wiederholten Ko-Registrierung. Dabei weist die GEE eine systematische Verschlechterung der Genauigkeit bei Variante „Pfad B“ auf, insbesondere in der y-Komponente. Diese Verschlechterung über die Zeitschritte geht bereits aus der Korrelationskonfidenz hervor (Tabelle 3), wobei 0 für eine niedrige und 1 für eine hohe Kon-

fidenz steht (GOOGLE EARTH ENGINE 2021). GeFolki liefert eine gleichbleibende Qualität. Insgesamt schneidet die „Pfad A“ Variante besser ab als die Ko-Registrierung bei gleichbleibender Referenz. Die Verschlechterungen bei „Pfad B“ sind auf die unterschiedliche Wolkenbedeckung zurückzuführen. Zudem entstehen durch die größeren Zeitabstände größere Veränderungen hinsichtlich Beleuchtung und spektralen Eigenschaften. Diese Unterschiede können in Variante „Pfad A“ besser berücksichtigt werden. Mit beiden Tools kann eine Subpixelgenauigkeit bei der Ko-Registrierung erreicht werden, wobei die verbleibenden Abweichungen kleiner als die ursprünglichen Differenzen in der Georeferenzierung sind.

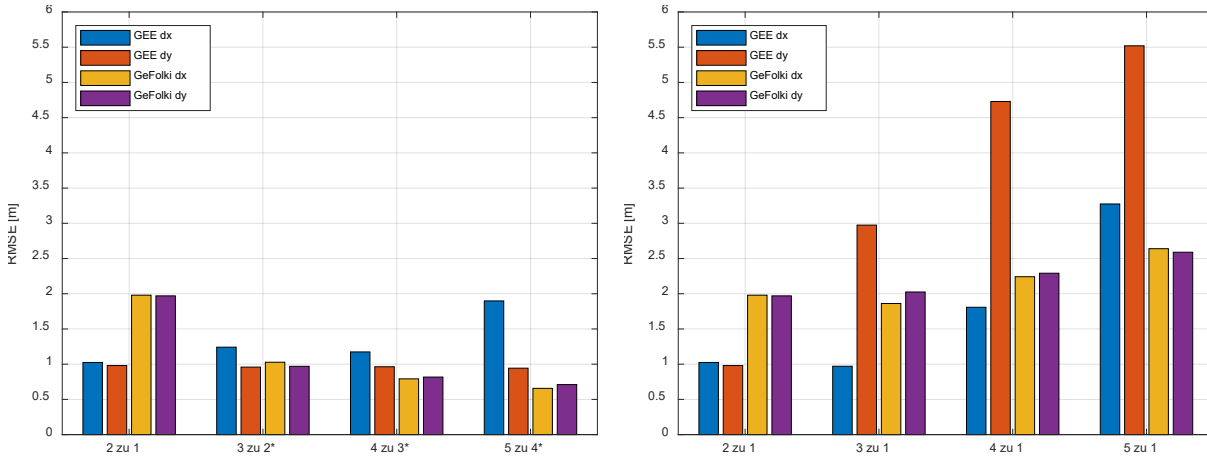


Abb. 2: Verbleibender Fehler nach zweimaliger Anwendung der Ko-Registrierung auf Sentinel-2 Zeitreihe bei „Pfad A“ (links) und „Pfad B“ (rechts).

Tab. 3: Durchschnittliche Korrelationskonfidenzen bei Verwendung der GEE für die Sentinel-2 Zeitreihe. Die Werte beziehen sich auf die Konfidenzen zwischen der Referenz und dem angegebenen Bild.

Bild	2	3	4	5
Pfad A	0,42	0,56	0,63	0,61
Pfad B	0,42	0,37	0,28	0,25

### 4.3 Sentinel-1 Zeitreihe

Im Gegensatz zu der Sentinel-2 Zeitreihe ist die bestimmte durchschnittlich notwendige Translation mit circa 5 m bei der GEE kleiner als bei GeFolki mit 7 m. Bezogen auf die Fehlerabschätzung schneidet GEE bei den Sentinel-1 Daten qualitativ besser als GeFolki ab (Abbildung 3). Beim Vergleich der Ko-Registrierungsstrategien kann keine eindeutige Präferenz festgestellt werden. Anhand der Korrelationskonfidenzen in Tabelle 4 zeigt sich eine etwas bessere Konfidenz für „Pfad A“. Jedoch verbleiben die Konfidenzen auf niedrigem Niveau. Mit Ausnahme des Ausreißers bei Bild 8 zu 1 sind jedoch die RMSE Werte in Abbildung 3 für „Pfad B“ besser als bei „Pfad A“. Im Gegensatz zur GEE und auch zur Sentinel-2 Ko-Registrierung können für GeFolki sowohl bei „Pfad A“ als auch „Pfad B“ systematische Unterschiede zwischen der x- und y-Komponente festgestellt werden, bei denen die y-Komponente schlechter abschneidet. Eine mögliche Ursache dafür kann in einem stärkeren Einfluss der Schrägsichtgeometrie bei den SAR Daten auf das GeFolki Verfahren liegen. Obwohl bei der Sentinel-1 Zeitreihe ebenfalls die Subpixelgenauigkeit erreicht

werden kann, ist festzustellen, dass im Vergleich zu der Sentinel-2 Zeitreihe die Qualität der eingesetzten Tools abnimmt. Zudem erreicht die verbleibende Abweichung in der Fehlerabschätzung in Einzelfällen das gleiche Maß wie die durchschnittlich bestimmte Ko-Registrierung selbst.

Tab. 4: Durchschnittliche Korrelationskonfidenzen bei Verwendung der GEE für Sentien-1 Zeitreihe.

	2	3	4	5	6	7	8	9	10	11
Pfad A	0,32	0,29	0,26	0,35	0,33	0,37	0,22	0,24	0,36	0,27
Pfad B	0,32	0,25	0,29	0,29	0,21	0,25	0,17	0,25	0,19	0,16

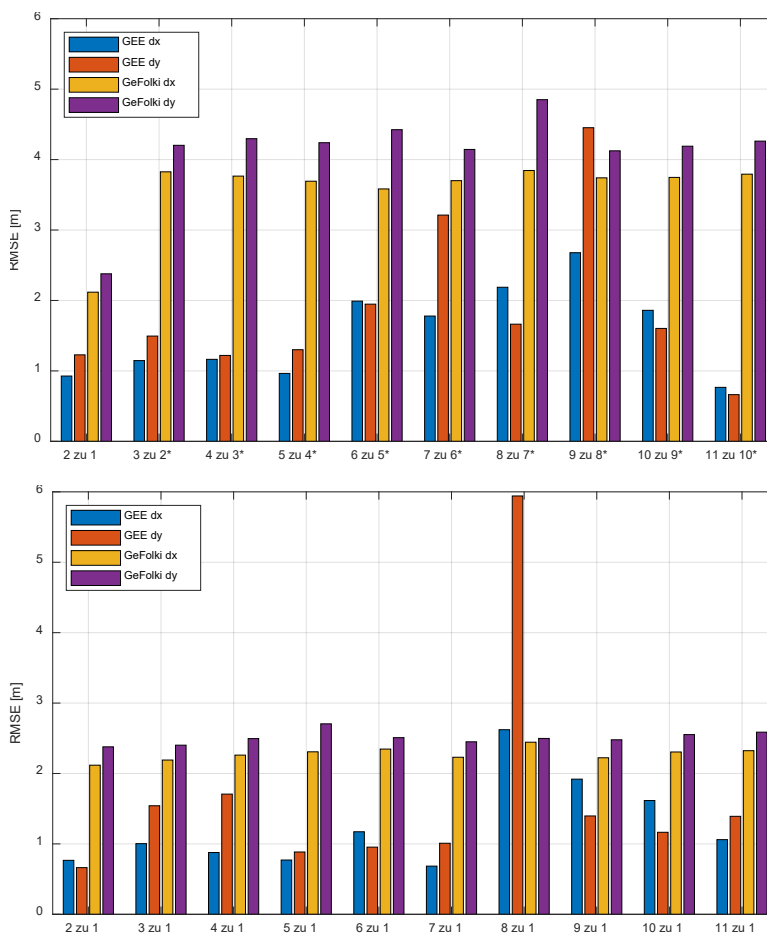


Abb. 3: Verbleibender Fehler nach zweimaliger Anwendung der Ko-Registrierung auf Sentinel-1 Zeitreihe bei „Pfad A“ (oben) und „Pfad B“ (unten).

#### 4.4 Multi-modale Ko-Registrierung

Für die multi-modale Ko-Registrierung mit SOMatch wurde bereits im Training des „Correspondence“ Netzwerkes ein durchschnittlicher Abstand zwischen Prädiktion und Referenz von 70 Pixeln erreicht. Dies deutet daraufhin, dass allein auf Basis von den Sentinel-1-2 Bildpaaren keine ausreichend korrelierenden Merkmale extrahiert werden konnten. Der erforderliche Grenzwert für eine erfolgreiche Zuordnung wurde somit nur von einem sehr kleinen Anteil der Trainingsdaten

erfüllt. Das hat zur Folge, dass keine Testdaten auf Basis von SOMatch ko-registriert werden konnten. Ein wesentlicher Faktor spielt dabei die Auslegung der Implementierung auf hochauflösende Satellitendaten oder alternativ die Hinzunahme von zusätzlichen Informationen wie der Landbedeckung. Somit ist trotz seiner Größe und entsprechender Anpassungen auf den SEN12 Datensatz die derzeitige Implementierung von SOMatch nicht für die Ko-Registrierung von Sentinel-1 mit Sentinel-2 Daten geeignet.

## 5 Fazit & Ausblick

Die Auswertung der verschiedenen Toolboxen zeigt, dass die bereits vorliegende Georeferenzierung der Datenprodukte eine Pixelgenauigkeit erreicht. Für die subpixel genaue Auswertung wird jedoch eine Ko-Registrierung benötigt. Insbesondere durch das gute Abschneiden bei der Sentinel-1 Ko-Registrierung wurde die GEE für die Datenvorverarbeitung der nachfolgenden Leckage Detektion eingesetzt. Der Einsatz von Deep Learning für den Ko-Registrierungsprozess benötigt jedoch in Zukunft eine robustere Trainierbarkeit, um effizient eingesetzt werden zu können. Zudem sind die in diesem Bereich eingesetzten punktbasierten Ansätze stark von der Auflösung und anderen datenvorverarbeitenden Schritten wie beispielsweise Speckle Filterung abhängig. Auf der Basis konventioneller Verfahren wie der GEE konnten jedoch vielversprechende Ergebnisse für die Leckage Detektion für Frei- und Landwirtschaftsflächen erzielt werden. Mit Hilfe der optischen Bildinformationen im Infrarotbereich können unter anderem Anomalien der Biomasse erkannt werden. Radardaten reagieren dagegen auf eine veränderte Dielektrizität des Bodens, die wiederum durch die Bodenfeuchte bestimmt wird. Beide Datenquellen zusammen liefern Hinweise auf Leckagen in den Wasserversorgungsleitungen. Diese entwickelte Technologie ermöglicht eine schnelle und genaue Identifikation von Leckagen in Versorgungsleitungen und damit die kostengünstige Überwachung von Netzabschnitte mit vorhandenen Satelliten.

## 6 Danksagung

Wir bedanken uns für die Projektförderung im Rahmen von LOEWE beim Land Hessen und der guten Zusammenarbeit mit der Spatial Business Integration GmbH. Zusätzlich möchten wir Jakob Bahle, Julian Schnell und Sebastian Hunger für ihre Unterstützung danken.

## 7 Literaturverzeichnis

- BOENING, F., RAKOVEC, O., KUMAR, R., SAMANIEGO, L., SCHRÖN, M., HILDEBRANDT, A., REBMANN, C., THOBER, S., MÜLLER, S., ZACHARIAS, S., BOGENA, H., SCHNEIDER, K., KIESE, R., ATTINGER, S. & MARX, A., 2022: High-resolution drought simulations and comparison to soil moisture observations in Germany. *Hydrol. Earth Syst. Sci.*, **26**, 5137-5161, <https://doi.org/10.5194/hess-26-5137-2022>.
- BRIGOT, G., COLIN-KOENIGUER, E., PLYER, A. & JANEZ, F., 2016: Adaption and Evaluation of an Optical Flow Method Applied to Coregistration of Forest Remote Sensing Images. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing*, **9**(7), 2923-2939, <https://doi.org/10.1109/JSTARS.2016.2578362>.

- GOOGLE EARTH ENGINE, 2021: Registering Images. <https://developers.google.com/earth-engine/guides/register>, letzter Zugriff am: 16.01.2023.
- HUGHES, L. H., MARCOS, D. LOBRY, S., TUIA, D. & SCHMITT, M., 2020: A deep learning framework for matching of SAR and optical imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, **169**, 166-179, <https://doi.org/10.1016/j.isprsjprs.2020.09.012>.
- KÄSTNER, S., 2023: Trockenheit im tiefen Boden behindert Grundwasserbildung. Deutschlandfunk, <https://www.deutschlandfunk.de/duerre-trotz-regen-warum-trockene-boeden-das-grundwasser-behindern-100.html>, letzter Zugriff am: 23.01.2023.
- KRAPEZ, J.-C., SANCHIS MUÑOZ, J., MAZEL, C., CHATELARD, C., DÉLIOT, P., FRÉDÉRIC, Y.-M., BARRILLOT, P., HÉLIAS, F., BARBA POLO, J., Olichon, V., SERRA, G., BRIGNOLLES, C., CARVALHO, A., CARREIRA, D., OLIVEIRA, A., ALVES, E., FORTUNATO, A.B., AZEVEDO, A., BENETAZZO, P., BERTONI, A. & LE GOFF, I., 2022: Multispectral Optical Remote Sensing for Water-Leak Detection. *Sensors*. **22**(3), 1057, <https://doi.org/10.3390/s22031057>.
- LAH, A. A. A., DZIYAUDDIN, R. A. & MD YUSOFF, N., 2018: Localization Techniques For Water Pipeline Leakages: A Review. 2nd International Conference on Telematics and Future Generation Networks (TAFGEN), Kuching, Malaysia, 49-54, <https://doi.org/10.1109/TAFGEN.2018.8580467>.
- SCHMITT, M., HUGHES, L. H. & ZHU, X. X., 2018: The SEN1-2 Dataset for Deep Learning in SAR-Optical Data Fusion. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, **IV-1**, 141–146, <https://doi.org/10.5194/isprs-annals-IV-1-141-2018>.
- SCHUSTER, K., 2022: Klimakrise: Wird in Deutschland das Trinkwasser knapp? <https://www.zdf.de/nachrichten/panorama/klima-wasser-knappheit-deutschland-100.html>, letzter Zugriff am 09.01.2023.
- TONDEWAD, P. S. & DALE, M. P., 2020: Remote Sensing Image Registration Methodology: Review and Discussion. *Procedia Computer Science*, **171**, 2390-2399, <https://doi.org/10.1016/j.procs.2020.04.259>.
- USS, M. L., VOZEL, B., LUKIN, V. V. & CHEHDI, K., 2016: Multimodal Remote Sensing Image Registration with Accuracy Estimation at Local and Global Scales. *IEEE Transactions on Geoscience and Remote Sensing*, **54**(11), 6587-6605, <https://doi.org/10.1109/TGRS.2016.2587321>.

# Punkt- und ebenenbasierte Detektion von Ecken und Kanten in Innenraum-Punktwolken

JAKOB SCHMIDT<sup>1</sup>, ANDREAS EICHHORN<sup>1</sup> & DOROTA IWASZCZUK<sup>2</sup>

*Zusammenfassung: In dieser Arbeit werden ein konturpunkt-basierter und ein ebenenbasierter Ansatz zur Extraktion von Kanten und Ecken aus Innenraumpunktwolken untersucht. Dabei liegt der Fokus auf der Detektion verdeckter Kanten und Ecken. Der ebenbasierte Ansatz erweist sich robuster gegenüber stark verrauschten und unvollständigen Punktwolken. Der punkt-basierte Ansatz erscheint durch die Verwendung lokaler Nachbarschaft zuverlässiger in komplexen Räumen. Beide Ansätze ermöglichen die Erfassung verdeckter Kanten. Der ebenbasierte Ansatz liefert dabei ein vollständigeres Ergebnis. Beide Ansätze erzeugen eine Übererfassung von Ecken und Kanten, daher ist eine Erweiterung der Ansätze zur Filterung und Verknüpfung der Ecken und Kanten sowie zur Darstellung als 3D-Modell notwendig.*

## 1 Einleitung

Zur Erfassung von Gebäudebeständen werden häufig Laserscanning und Photogrammetrie eingesetzt. Dabei werden große Datenmengen mit vielen Punkten erzeugt. Diese sind in der Regel unstrukturiert und werden als Basis für die Modellierung z. B. eines Building Information Models (BIM) verwendet (LÓPEZ IGLESIAS et al. 2020). Im Gegensatz dazu werden bei einer klassischen tachymetrischen Objektvermessung Ecken direkt gemessen. Die Oberflächen werden dabei auf die minimal notwendigen Punkte beschränkt. Die Ecken werden bei der Messung durch den Anwender manuell in Beziehung zueinander und zu den beschriebenen Objekten gesetzt. Um vergleichbare Informationen aus photogrammetrischen und Laserscanning-Punktwolken zu gewinnen, ist es notwendig Informationen über Objekte und deren Beziehungen zueinander zu extrahieren. Im Folgenden werden zwei Ansätze zur Extraktion von Ecken und Kanten aus Punktwolken von Innenräumen betrachtet und verglichen. Dabei basiert eines auf der Detektion von Ebenen (SCHNABEL et al. 2007) und das andere auf der Detektion von Konturpunkten (AHMED et al. 2018).

## 2 Problemstellung

Die Erfassung von Innenräumen über photogrammetrische Verfahren und Laserscanning ermöglicht eine schnelle und vollständige Erfassung von Gebäuden in Form von Punktwolken. Aus diesen Daten Innenraummodellen auf manuelle Weise zu erzeugen ist sehr arbeits- und zeitaufwändig. Deshalb ist die automatische Extraktion von Ecken und Kanten zur Verarbeitung von Punktwolken essenziell für die Erfassung von Innenräumen als 3D-Modelle (LÓPEZ IGLESIAS et al. 2020).

---

<sup>1</sup> Technische Universität Darmstadt, Fachgebiet Geodätische Messsysteme und Sensorik, Franziska-Braun-Straße 7, D-64287 Darmstadt, E-Mail: Jakob.Schmidt1@tu-darmstadt.de, eichhorn@geod.tu-darmstadt.de

<sup>2</sup> Technische Universität Darmstadt, Fachgebiet Fernerkundung und Bildanalyse, Franziska-Braun-Straße 7, D-64287 Darmstadt, E-Mail: Dorota.Iwaszczuk@tu-darmstadt.de

Dieser Arbeit liegt die Idee zugrunde, Ecken und Kanten in Innenraum-Punktwolken zu erfassen und die Beziehungen dazwischen zu bestimmen. Dabei soll eine polygonisierte Darstellung von Innenräumen mit einem Ansatz, der es ermöglicht, verdeckte Kanten und Ecken zu erfassen, umgesetzt werden. In aktuellen Publikationen wird dafür unter anderem Deep Learning verwendet. Abhängig vom Verfahren wird dabei davon ausgegangen, dass Ecken im Raum in einer Punktwolke repräsentiert sind, wodurch Verdeckungen nicht berücksichtigt werden können (LIU et al. 2021). Passende Trainingsdatensätze mit verdeckten Kanten, um ein entsprechendes Netz zu trainieren, sind den Autoren bisher nicht bekannt. Die Beschaffung solcher Trainingsdaten wäre zeitaufwändig. Um diesem Problem auf andere Weise zu begegnen, werden in dieser Arbeit zwei geometrische Ansätze zur Extraktion von Ecken und Kanten aus Innenraum-Punktwolken betrachtet und verglichen. Diese sind grundsätzlich geeignet, um verdeckte Kanten zu bestimmen und benötigen keine Trainingsdaten. Der erste Ansatz basiert auf der Detektion von Konturpunkten, der zweite auf der Detektion von Ebenen. Der erste wird im Folgenden als punktbasiert und der zweite als ebenbasierter Ansatz bezeichnet. Der Fokus der polygonisierten Darstellung der Innenraumgeometrie liegt dabei auf Wänden, Decken und Böden. Dabei wird untersucht, welcher Ansatz sich besser eignet und welche Vor- und Nachteile in Bezug auf Vollständigkeit und Korrektheit bestehen. Diese Untersuchung dient als Grundlage für die Entwicklung eines Konzepts zur Erfassung von Innenräumen, welches kein Deep Learning benötigt.

### 3 Verwandte Arbeiten

Für die Extraktion dreidimensionaler Kanten in Punktwolken existieren unterschiedliche Lösungsansätze. Dazu gehört die Verwendung von Ebenen und deren Schnittstellen, das direkte Klassifizieren von Kanten, semantische Segmentierung von Ebenen und letztlich Deep Learning zum Lernen und Detektieren von Ecken und Kanten.

Um die Konturen von Gebäuden in Punktwolken zu bestimmen, nutzen LU et al. (2019) eine zweidimensionale Kantendetektion. Dafür werden 3D-Punkte von Ebenen in eine 2D-Darstellung übertragen. Diese Darstellung wird zur Kantendetektion binarisiert. Die resultierenden Kanten werden in den dreidimensionalen Raum zurücktransformiert. Um die Konturen direkt aus Punktwolken zu bestimmen, können Punkte entlang von Ecken und Kanten klassifiziert werden. HACKEL et al. (2016) nutzen dafür geometrische Merkmale basierend auf Eigenwerten und Eigenvektoren. Für jeden Punkt wird auf Basis dieser Merkmale die Wahrscheinlichkeit für einen Konturpunkt abgeschätzt. Aus gleichmäßig verteilten Punkten, die einer Kantenbedingung entsprechen, wird ein Graph aufgebaut, der Kantenkandidaten beinhaltet. In diesem Graph wird aus Nachbarschaftsbeziehungen eine Untermenge an Punkten gewählt, die der Kontur auf Basis geometrischer Merkmale am besten entsprechen. Auf vergleichbare Weise nutzen AHMED et al. (2018) die Symmetrie der lokalen Nachbarschaft, um Punkte entlang von Kanten zu extrahieren und diese anschließend mit Hilfe von Krümmungsvektoren in Eck- und Konturpunkte zu trennen. Dieser Ansatz wird in dieser Arbeit als Grundlage verwendet. Andere Publikationen verwenden neuronale Netze zur Bestimmung von Konturpunkten. HIMEUR et al. (2020) beschreiben dafür Punkte und deren lokale Nachbarschaft in unterschiedlicher Skalierung. Es werden die Rauheit, die Normalen und die Krümmung betrachtet. Punkte, die direkt auf einer Kante liegen, werden als scharfe Kante

definiert und Punkte in der unmittelbaren Umgebung als weiche Kante. Damit wird ein neuronales Netz zur Kantendetektion trainiert. Da nur wenige Punkte nah an einer Kante zu liegen kommen, erhöht die Einbeziehung der umgebenden Punkte, gemäß der Autoren, die Robustheit der Detektion. Dieser Ansatz benötigt nur wenige manuell annotierte Trainingsdaten und ermöglicht eine Anpassung auf die spezifischen Eigenschaften einer Punktwolke, die z. B. durch unterschiedliche Erfassungsmethoden entstehen. Der klassische Ansatz zur Bestimmung der Geometrie von Innenräumen ist die Detektion von Wänden, Böden und Decken über Ebenen (VILLE et al. 2021). Zur Bestimmung dieser Ebenen können z. B. Local Hough Voting oder RANSAC verwendet werden (SOMMER et al. 2020). Sie bestimmen orthogonale Ebenen in einer Punktwolke und verschneiden diese, um Schnittgeraden aus zwei Ebenen entlang von Kanten und Schnittpunkte aus drei Ebenen als Eckpunkte zu berechnen. Eine semantische Segmentierung ermöglicht die Erweiterung dieses Ebenenansatzes (CASTAGNO & ATKINS 2020). verwenden diese zur Erfassung von Strukturen in Innenräumen und zur Repräsentation als Polygone. Dazu werden Meshs aus Punktwolken mit Hilfe der Normalenvektoren segmentiert und zu Flächen mit gleicher Ausrichtung zusammengefasst. Anschließend werden Polygone bestimmt, die diese Flächen umschließen, um die Strukturen in Innenräumen zu repräsentieren. IWASZCZUK et al. (2017) nutzen semantische Informationen über Konturen und Ebenen aus Bildern, um diese auf Tiefenkarten zu übertragen. Damit erzeugen sie eine semantisch angereicherte Punktwolke, aus der Ebenen, mittels k-Means-Clustering und RANSAC, bestimmt werden. Ein Verfahren, das ohne die Erfassung von Ebenen funktioniert, ist die direkte Erstellung von Drahtgittermodellen aus Punktwolken mittels Deep Learning (LIU et al. 2021). Die Autoren stellen ein Feed-Forward-Neuronal-Network vor, das für die Anwendung auf CAD-Modelle von einzelnen Objekten konzipiert wurde. Über mehrere Netze werden zuerst Eckpunkte bestimmt. Diese Eckpunkte werden anschließend mit Kanten verknüpft.

## 4 Methodik

Im folgenden Kapitel wird die Vorgehensweise der punktbasierten und ebenbasierten Extraktion von Kanten und Ecken aus Innenraum-Punktwolken beschrieben.

### 4.1 Punktbasierter Ansatz

Beim punktbasierten Ansatz werden in einem ersten Schritt Konturpunkte extrahiert. Dafür wird die Methode von AHMED et al. (2018) verwendet. Zu jedem Punkt wird die Symmetrie der lokalen Nachbarschaft betrachtet und ein Schwerpunkt bestimmt. Bei großer Abweichung des Punkts zum Schwerpunkt kann von einem Eck- oder Kantenpunkt ausgegangen werden. Anschließend werden Krümmungsvektoren in Form der Eigenvektoren aus den kleinsten Eigenwerten verwendet, um die Punkte in Eck- und Konturpunkte zu klassifizieren. Abhängig von der Form der Punktwolke und der Punktzahl ist diese Vorgehensweise sehr rechenintensiv, daher wird ein Subsampling auf einen Punkt pro Zentimeter der Originalpunktwolke durchgeführt, um die Rechenzeit und den Speicherbedarf zu reduzieren. Dieser Ansatz wird mittels Random Sample Consensus (RANSAC) erweitert, um Geraden in die Konturpunkte zu legen. Durch RANSAC wird eine zufällige Stichprobe gewählt und geprüft, wie gut die umgebenden Punkte zum vorgegebenen Modell einer Gerade passen. Da diese Geraden unterschiedlich lang sein können und unterschiedlich viele Punkte beinhalten, hängt das Ergebnis der Detektion stark von der Wahl der Parameter für



RANSAC ab. Deshalb wird die maximale Distanz für Inlier, die Anzahl an Iterationen und die Anzahl an Inliern für eine Gerade in dieser Arbeit empirisch festgelegt. Zuerst werden 500 Geraden geschätzt und davon die Gerade mit den meisten Inliern, mit einer maximalen Distanz von 5 cm, gewählt. Die zugehörigen Punkte werden aus der Punktwolke entfernt und von der weiteren Bestimmung ausgeschlossen. Dieser Prozess wird wiederholt, bis 100 Geraden gefunden wurden. Diese Vorgehensweise ist notwendig, um mögliche Geraden zu finden, hat jedoch den Nachteil, dass die Eckpunkte, als Schnittpunkte mehrerer Geraden, nach der ersten Zuordnung zu einer Gerade für die Bestimmung weiterer Geraden nicht mehr zur Verfügung stehen. Um diesem Problem und der unterschiedlichen Anzahl an Punkten pro Kante zu begegnen und alle Kanten zu finden, wird die anfangs geforderte Anzahl von 210 Inliern bei jeder Iteration um 2 reduziert. Diese Geraden werden als Kantenkandidaten interpretiert. Unter der Annahme von Manhattan-World (COUGHLAN & YUILLE 1999) werden aus diesen Kantenkandidaten nur solche gewählt, die nahezu parallel zu den Koordinatenachsen verlaufen. Jeweils aus zwei Geraden wird ein Schnittpunkt berechnet, der als Eckpunkt interpretiert wird. Aufgrund von Messungenauigkeiten ist davon auszugehen, dass die Geraden entlang der Kanten windschief verlaufen. Daher wird ein Pseudo-Schnittpunkt als Mittelpunkt der geringsten Distanz zweier Geraden als Eckpunktkandidat berechnet. Durch die Wahl eines maximal erlaubten minimalen Abstands zwischen Geraden wird ermittelt, welche Kanten potenziell einen gemeinsamen Schnittpunkt besitzen. Die resultierenden Schnittpunkte werden über eine Bounding Box um die Punktwolke gefiltert, um Ausreißer zu entfernen. Da sich in einem Eckpunkt grundsätzlich drei Geraden treffen, wird zu diesen Schnittpunkten die geringste Distanz zu allen anderen Kanten bestimmt. Liegt eine dicke Kante in einer maximalen Entfernung von 20 cm zu einem Eckpunkt, wird der zugehörige Eckpunkte bestätigt und als Mittelpunkt zu dieser Gerade neu bestimmt. Somit können theoretisch auch verdeckte Eckpunkte erfasst werden.

## 4.2 Ebenenbasierter Ansatz

Der ebenenbasierte Ansatz entspricht der Vorgehensweise von SOMMER et al. (2020). Mit Hilfe von M-Estimator Sample Consensus (MSAC) werden Ebenen in die Punktwolke gelegt, diese können als Kandidaten für Wandelemente betrachtet werden. MSAC ist eine Erweiterung von RANSAC, bei dem die Inlier abhängig zur Distanz zum gewählten Modell gewichtet werden. MSAC wird aufgrund der Robustheit und einfachen Implementierung genutzt. Dabei wird ein maximaler Punktabstand zur Ebene von 5 cm verwendet. Es werden Ebenen detektiert und die zugehörigen Punkte entfernt, bis keine Ebenen mit mehr als 1000 Punkten gefunden werden können. Die Verschneidung zweier Ebenen erlaubt die Berechnung von Schnittgeraden. Diese entsprechen den Kantenkandidaten. Die Schnittbedingung wird mit Hilfe des Kreuzprodukts der Normalenvektoren unter Berücksichtigung einer Toleranz auf nahezu rechtwinklige Schnitte eingeschränkt. Ein Schnittpunkt von drei Ebenen entspricht einem potenziellen Eckpunkt. Diese Vorgehensweise hat den Vorteil, dass ein Eckpunkt immer auf einem Geradenschnittpunkt von drei Geraden liegt. Die Schnittgeraden und Schnittpunkte, die durch Messfehler außerhalb des Raums liegen, werden außerhalb einer Bounding Box abgeschnitten, die die gesamte Punktwolke mit einer Toleranz von 20 cm umschließt. Zusätzlich wird eine Klassifikation implementiert, die Punkte und Geraden als weniger wahrscheinlich einstuft, die mit einem Toleranzmaß von 5 cm

außerhalb der Punktwolke liegen. Zu jedem Eckpunkt wird dazu die minimale Distanz zur Punktwolke bestimmt. Da verdeckte Ecken ebenfalls eine große Entfernung zur Punktwolke aufweisen eignet sich diese Klassifikation nicht zum eindeutigen Ausschluss von Eckpunkten und Kanten.

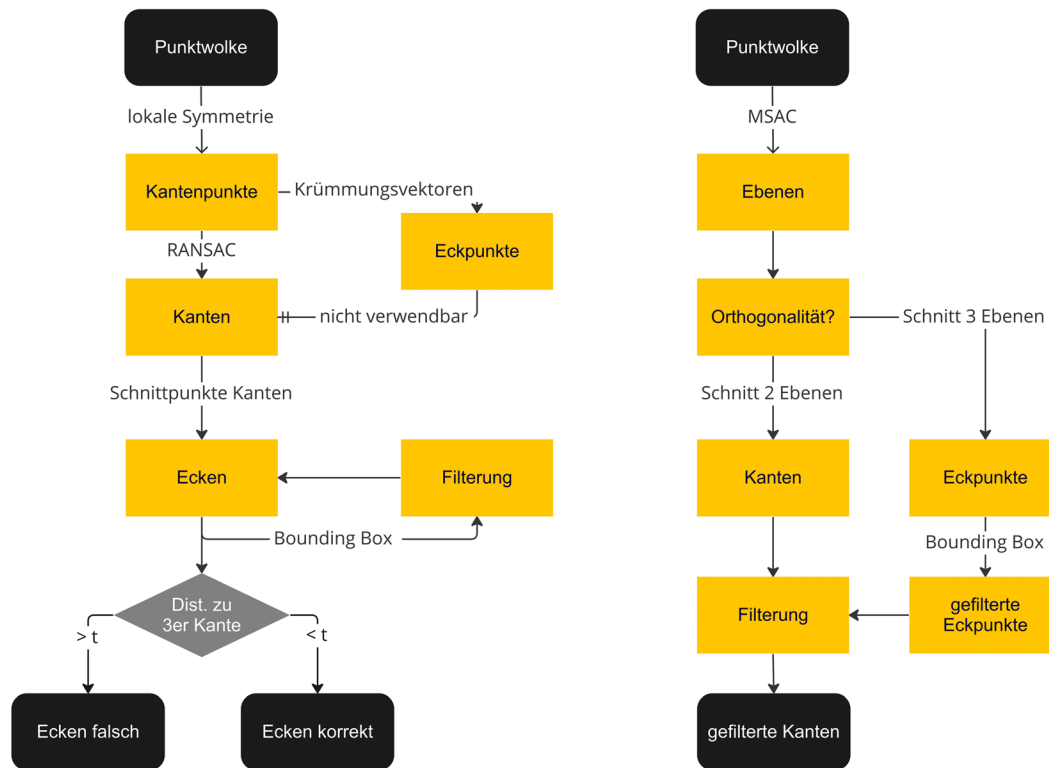


Abb. 1: Ablaufdiagramm punktbasierter (links) und ebenbasierter Ansatz (rechts)

## 5 Experimente

Die vorgestellte Methodik wird anhand von zwei Datensätzen getestet (siehe Tab. 1 und Abb. 2), dem „ISPRS Benchmark on Indoor Modelling“, im Folgenden BoIM genannt, (KHOSHELHAM et al. 2018) und dem „Stanford large-scale 3D Indoor Spaces Dataset“ (S3DIS) (ARMENI et al. 2016). Die Datensätze bestehen aus Punktwolken von Stockwerken mehrerer Gebäude. Der S3DIS-Datensatz wurde mit einer Matterport-Kamera erfasst, die auf Basis von Structured Light Tiefeninformationen liefert und daraus eine Punktwolke erzeugt. Der BoIM-Datensatz wurde mit einem Mobile Laserscanning System erfasst. Dieses Trolley-System besteht aus drei einzelnen Laserscannern und einer 360°-RGB-Kamera für die Farbinformationen. Die verwendete Punktwolke beinhaltet jedoch keine Farbinformation. Abb. 2 macht deutlich, dass der S3DIS-Datensatz eine deutlich gleichmäßigere Punktverteilung aufweist als der BoIM-Datensatz. Der S3DIS-Datensatz beinhaltet in jedem Raum das Mobiliar, bei BoIM existieren Stockwerke und Räume mit und ohne Mobiliar. Als geometrische Referenz stehen bei BoIM von Hand modellierte BIM-Daten zur Verfügung. Bei S3DIS existiert nur eine semantisch gelabelte Punktwolke. Das BIM eignet sich nur eingeschränkt zur Kontrolle der Ergebnisse, denn ein BIM besteht aus volumetrischen

Wandelementen. Im Vergleich dazu korreliert eine Indoor-Punktwolke stärker mit der Repräsentation von Wänden als unendlich dünne Ebenen (VILLE et al. 2021). Dadurch entstehen Unterschiede sowohl an Ecken als auch an Kanten zwischen der BIM-Referenz und der Punktwolke. Bei einem BIM sind z. B. die Kanten, an denen sich die Wände im Innenraum berühren, nicht für alle Wände identisch modelliert. Deshalb lassen sich die Innenraum-Kanten und Eckpunkte in Position und Anzahl nicht direkt aus dem BIM extrahieren und vergleichen. Da die Anpassung der Referenz zur Kontrolle der Ergebnisse nicht ohne Weiteres möglich ist, wurden die detektierten Ecken und Kanten vorerst nur visuell auf Plausibilität geprüft. Dafür wurden die Kanten und Eckenzahl pro Raum in der Originalpunktwolke gezählt und mit der Anzahl an Ecken und Kanten aus den beiden Ansätzen verglichen (Tab. 2). Die vorliegende Betrachtung erfolgt für einzelne Räume und nicht für gesamte gesamten Stockwerke. Aus dem S3DIS-Datensatz wird ein Konferenzraum mit Mobiliar verwendet, aus dem BoIM-Datensatz ein Raum ohne Mobiliar. Aus dem S3DIS-Datensatz wurde der „conferenceRoom\_1“ in „Area\_1“ mit einer Grundfläche von 4×5 m, einer Höhe von 3 m und einer Punktzahl von 1,95 Mio. verwendet. Aus dem BoIM-Datensatz wurde ein Raum aus „CaseStudy1\_TUB1“ mit einer Fläche von 5×7 m, einer Höhe von ca. 2,5 m und einer Punktzahl von 1,14 Mio. ausgewählt (Tab. 1). Auf die beiden Datensätze werden jeweils der punktbasierte und der ebenenbasierte Ansatz angewendet. Der Vergleich der Ergebnisse der beiden Datensätze lässt zum einen Rückschlüsse auf die Auswirkung der Erfassungsmethoden auf die Kanten- und Eckendetektion zu, zum anderen lässt sich durch die unterschiedliche Möblierung der Räume auch eine Einschätzung der Ansätze in Bezug auf die Erfassung verdeckten Ecken und Kanten ableiten.

Tab. 1: Parameter der untersuchten Datensätze

	<b>BoIM</b>	<b>S3DIS</b>
<b>Raum</b>	Raum aus CaseStudy1_TUB1	conferenceRoom_1 in Area_1
<b>Größe</b>	5x7x2,5 m	4x5x3 m
<b>Sensor</b>	Viametris iMS3D	Matterport Kamera
<b>Punktzahl</b>	1,14 Mio.	1,95 Mio.
<b>Referenz</b>	BIM-Modelle	semantisch gelabelte Punktwolke

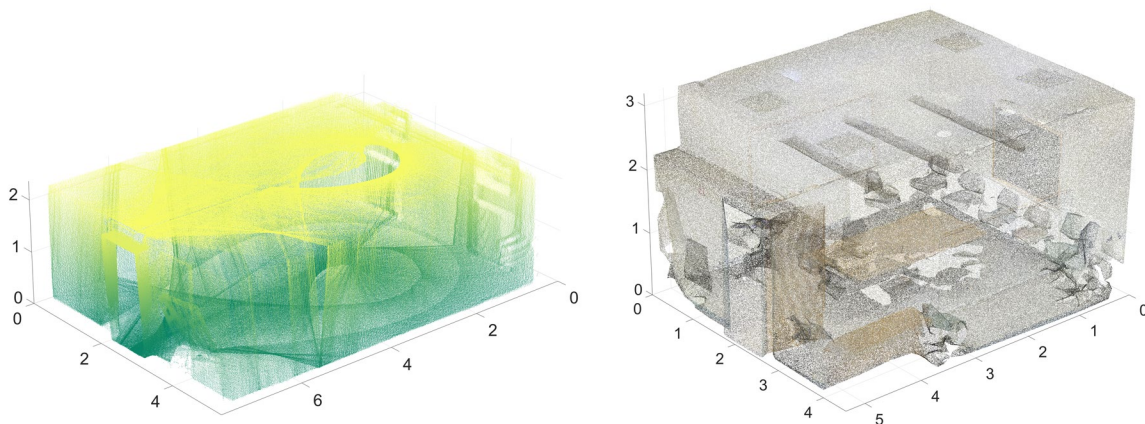


Abb. 2: Originalpunktwolke BoIM (links) und S3DIS (rechts) [m]

## 6 Ergebnisse

In Tab. 2 wird die tatsächliche und die durch die vorgestellten Ansätze erfasste Anzahl an Ecken und Kanten gegenübergestellt. Kanten und Ecken wurden manuell ohne Berücksichtigung von Mehrfachbestimmung gezählt. Die Anzahl an Ecken- und Kantenkandidaten stellt die tatsächlich vom Algorithmus bestimmte Anzahl der Ecken und Kanten dar.

Tab. 2: Tatsächliche und bestimmte Anzahl an Ecken und Kanten mit beiden Datensätzen und Ansätzen

	tatsächlich		punktbasiert		ebenenbasiert	
	BoIM	S3DIS	BoIM	S3DIS	BoIM	S3DIS
<b>Kanten</b>	12	18	10	12	12	17
<b>Ecken</b>	8	12	7	4	8	11
<b>verdeckte Ecken</b>	0	4	0	2	0	4
<b>Kantenkandidaten</b>	-	-	49	51	93	180
<b>Eckkandidaten</b>	-	-	141	92	188	441

### 6.1 Punktbasierter Ansatz

Der punktbaasierte Ansatz liefert mit S3DIS 12 von 18 Kanten und 7 von 12 Ecken (siehe Tab. 2). Von den 4 verdeckten Ecken wurden 2 gefunden, mit BoIM sind es 10 von 12 Kanten und 6 von 8 Ecken. Im S3DIS-Datensatz wurden 51 Kanten- und 92 Eckkandidaten gefunden, mit BoIM 49 Kanten- und 141 Eckkandidaten. Abb. 3 zeigt die Extrahierte der Konturpunktwolke Dabei werden mit S3DIS ca. 31500 Punkte und mit BoIM ca. 16700 Punkte als Konturpunkte bestimmt. Es ist zu sehen, dass sich die ungleichmäßige Punktwolke bei BoIM auch zu einer weniger vollständigen und stärker verrauschten Konturpunkt wolke führt als bei S3DIS. Die Konturen der Objekte im Raum können bei S3DIS ebenfalls mitbestimmt werden. Es lässt sich beobachten, dass mit BoIM ebenfalls Punkte als Konturpunkte eingestuft werden, die durch Scanmuster entstanden sind und nicht auf einer Kante liegen. Diese Punkte bilden kurvenförmige Strukturen. Für den punktbaasierten Ansatz ist neben der vollständigen Konturerfassung besonders die geradlinige Erfassung der Kanten für alle folgenden Schritte wichtig. Das Messverfahren mittels MLS führt bei BoIM zu unvollständigen und falsch bestimmten Konturen. Auch die Raummessung mit Structured Light bei S3DIS erzeugt Konturen, die stark von Geraden abweichen. Daneben ist auffällig, dass besonders an der Tür bei S3DIS sehr viele Kanten und Eckpunkte falsch bestimmt werden, in Abb.2 scheint diese ebenfalls ungleichmäßig erfasst zu sein.

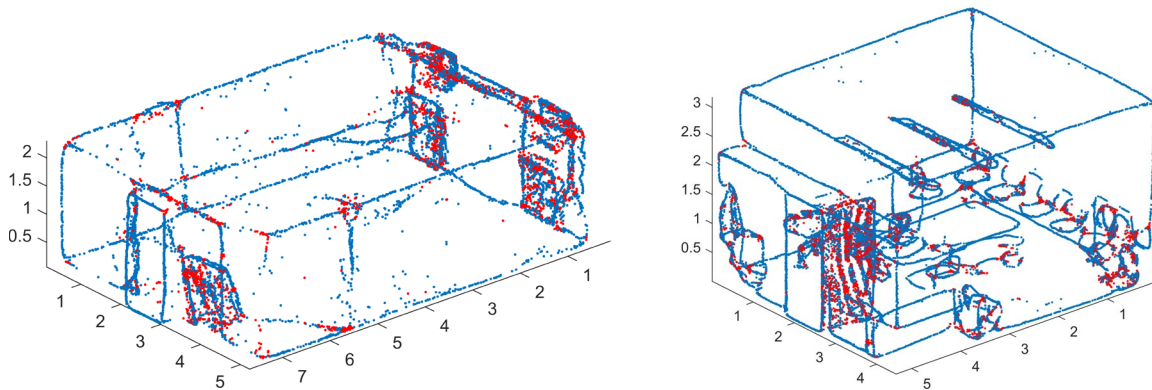


Abb. 3: Kantenpunktwolke (blau) mit Eckpunkten (rot) aus Krümmungsvektoren (links BoIM, rechts S3DIS) [m]

Die Weiterverarbeitung der Konturpunkte durch die Trennung von Eck- und Konturpunkten mithilfe von Krümmungsvektoren erweist sich als sehr unzuverlässig (Abb. 3). Es werden sehr viele Punkte falsch als Eckpunkte bestimmt. Diese werden daher nicht weiterverwendet. Stattdessen ermöglicht die Bestimmung der Pseudo-Schnittpunkte von drei windschiefen Geraden entlang der Kanten die Erfassung verdeckter Eckpunkte (Abb. 3). Die Ungenauigkeiten in der Konturpunktwolke führen dazu, dass an den betreffenden Kanten mittels RANSAC falsche oder keine Geraden gefunden werden. Die Filterung der Kanten unter der Manhattan-World-Annahme reduziert diese falschen Beobachtungen teilweise, aber nicht vollständig. Die Bestimmung von Pseudo-Schnittpunkten aus diesen Geraden erzeugt eine unvollständige und fehlerhafte Erfassung der Eckpunkte. Das Mobiliar im Raum im S3DIS-Datensatz führt ebenfalls zu falsch erkannten Kanten und Ecken. Der Schnitt von zwei Geraden beim punktbasieren Ansatz ermöglicht ebenfalls die Erfassung von regelmäßigen Öffnungen in den Wänden, wie z. B. des Türrahmens bei BoIM (Abb. 4).

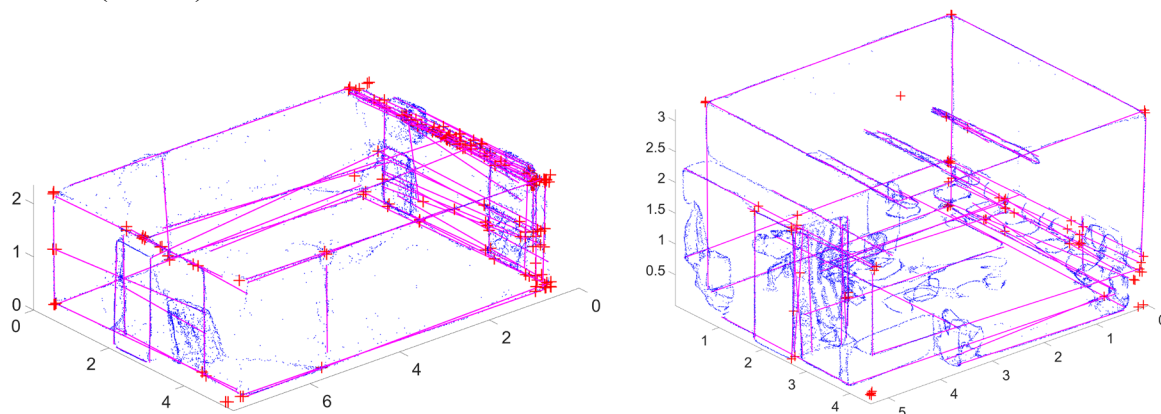


Abb. 4: Kantenpunktwolke (blau), Kanten- (magenta) und Eckkandidaten (rot) (links BoIM, rechts S3DIS) [m]

Da sich am Türrahmen in der Regel jedoch keine drei Kanten schneiden, erfüllen diese die Eckbedingung nur teilweise. Der Vergleich zwischen S3DIS und BoIM deutet darauf hin, dass eine gleichmäßige Punktverteilung auf den Wänden sowie entlang der Kanten- und Eckpunkte entscheidend für eine korrekte Extraktion von Ecken und Kanten ist. Dies gilt sowohl für die

Konturextraktion, als auch für die Bestimmung der Kanten mittels RANSAC. Die Grundform des Raums wird mit der punktbasierten Methode insgesamt unvollständig erfasst. Aufgrund von nicht detektierte Kanten werden die Ecken nur teilweise erfasst.

## 6.2 Ebenenbasierter Ansatz

Mit dem ebenbasierten Ansatz können mit dem S3DIS-Datensatz 17 von 18 Ecken- und 11 von 12 Kanten erfasst werden. Die 4 verdeckten Kanten wurden vollständig erfasst (Tab. 2). Mit BoIM wurden alle 8 Ecken und 12 Kanten vollständig erfasst. Insgesamt wurden bei S3DIS 180 Kanten- und 441 Eckkandidaten gefunden. Für BoIM sind es 93 Kanten- und 188 Eckkandidaten. Die Ebenen der Wände können mittels MSAC vollständig bestimmt werden (Abb. 5). Die Verschneidung dieser Ebenen führt zur Detektion vieler Geraden, die sich großteils im Bereich der tatsächlichen Kanten befinden. Diese Kantenkandidaten sind bündelartig angeordnet. An den Wänden, der Decke und dem Boden entspricht dabei an den meisten Kanten mindestens eine Gerade näherungsweise der tatsächlichen Kante. An den schrägen Schnittgeraden an der Decke (Abb. 5 BoIM) ist zu erkennen, dass die Ebenen bei kurzen Wandabschnitten, z. B. im Türbereich nicht parallel zur eigentlichen Wand verlaufen. Die Objekte im Raum besonders beim S3DIS-Datensatz genügen teilweise ebenfalls der Ebenenbedingung, daher werden viele falsche Ebenen detektiert. Dies führt wiederum zu falschen Kanten- und Eckkandidaten. Die orthogonale Schnittbedingung reduziert die Anzahl falsch bestimmten Ecken und Kante deutlich, bedingt bei S3DIS jedoch auch, dass nicht alle Kanten im Raum vollständig erfasst werden. Dies liegt vermutlich daran, dass die Wände im S3DIS-Datensatz weniger plan sind, als bei BoIM. Somit treten hier verstärkt unvollständig erfasste Ecken und Kanten auf. Die verdeckten Ecken wurden mit dem ebenbasierten Ansatz vollständig bestimmt (Tab. 2). Die Grundform des Raumes wurde nahezu vollständig erfasst, es tritt jedoch eine deutliche Übererfassung der Ecken und Kanten auf.

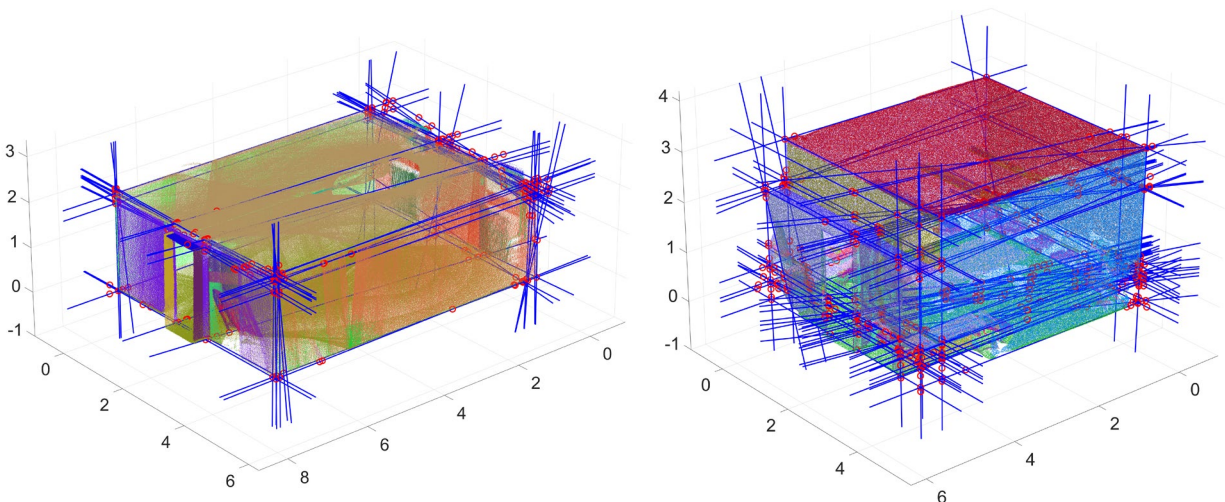


Abb. 5: Unterschiedlich gefärbte Ebenen mit Kanten- (blau) und Eckkandidaten (rote Kreise) (links BoIM, rechts S3DIS) [m]

## 7 Diskussion

Die Bestimmung von Ebenen erweist sich in den untersuchten Datensätzen als grundsätzlich robuster, als die Bestimmung von Konturen. Dies liegt vermutlich auch daran, dass für eine korrekte Erfassung einer Wand nur ein Teil sichtbar sein muss. Daneben ist die Wahrscheinlichkeit, dass eine Kante z. B. durch Möbel komplett verdeckt ist deutlich höher, als dass eine gesamte Wand nicht sichtbar ist. Die zuverlässigere Ebenendetektion führt ebenfalls zu einer vollständiger Kanten- und Eckenerfassung. Dies liegt vermutlich auch daran, dass eine Ebene Informationen über mehrere sowohl sichtbare als auch verdeckte Ecken und Kanten liefert. Eine Kante beinhaltet dagegen nur Informationen über zwei Eckpunkte. In Abb. 4 und 5 ist bei BoIM an der rechten Seite des Raums zu erkennen, dass ungenau erfasste Fenster in der Punktwolke mit beiden Ansätze zu einer großen Anzahl falsch bestimmter Kanten und Ecken führen. Dies zeigt, dass eine ungenaue Punktwolke die Auswahl korrekter Kanten und Ecken deutlich erschwert. Die Vollständige Erfassung der Ecken und Kanten bei BoIM zeigt ebenfalls, dass die Komplexität des Raums für die vollständige und korrekte Bestimmung aller Wände durch Ebenen entscheidend ist. Die stark variierenden Eigenschaften der einzelnen Wände führen demnach zu Unter- oder Überanpassung. Diese Übererfassung komplexer Räume, zeigt sich vermutlich bei S3DIS im Vergleich zu BoIM durch die deutlich größeren Anzahl an Kanten- und Eckkandidaten bei Verwendung des ebenbasierten im Vergleich zum punktbasierten Ansatz (Tab. 2). Bei BoIM unterscheiden sich die Ansätze in der Kanten- und Eckkandidatenzahl nur wenig. Die Komplexität des Raums hat durch die Verwendung der lokalen Nachbarschaft scheinbar nur einen geringen Einfluss. Die vorhandenen Unterschiede zeigen, dass die vollständige Erfassung aller Kanten mit dem punktbasierten Ansatz durch die Schätzung von Geraden aus der Kantenpunktwolke grundsätzlich ebenfalls von der variierenden Punktzahl und Dichte der Punkte entlang von Kanten unterschiedlicher Länge abhängig ist. Da alle Ecken und Kanten mit dem ebenbasierten Ansatz unter Verwendung des BoIM-Datensatz vollständig erfasst werden, kann davon ausgegangen werden, dass bei Verwendung von MSAC eine gleichmäßige Punktverteilung nicht notwendig ist, um Wandflächen zuverlässig zu erkennen. Die punktbasierte Methode erscheint deutlich abhängiger von gleichmäßig erfassten Kanten und Ebenen, wodurch z. B. nicht alle verdeckten Ecken gefunden werden konnten.

Neben der Zuverlässigkeit bei der Bestimmung von Ecken und Kanten unterscheiden sich die Ansätze ebenfalls im rekonstruierbaren Modell. So ermöglicht der ebenbasierte Ansatz die volumetrische Rekonstruktion von Innenräumen, durch die Erfassung der Wände als Flächen. Der punktbasierte Ansatz erzeugt nur ein Drahtgittermodell. Um vergleichbare Modelle aus beiden Ansätzen zu erhalten, müssen aus dem Drahtgittermodell zusätzlich Flächen aus umschließenden Kanten und Ecken bestimmt werden.

## 8 Fazit & Ausblick

Sowohl mit dem ebenbasierten als auch mit dem punktbasierten Ansatz ist die Erfassung der Geometrie eines Raumes mit Hilfe von Ecken und Kanten aus Punktwolken möglich. Beide Ansätze eignen sich zudem zur Detektion verdeckter Ecken und Kanten. Dabei stellt das Herausfiltern

falsch bestimmter Ecken und Kanten die größte Herausforderung dar. Die Bestimmung der Kanten und Eckpunkte mit Hilfe von Ebenen erweist sich als zuverlässiger und robuster als über Konturpunkte, insbesondere bei ungleichmäßig erfassten Punktwolken. Dabei muss eine sinnvolle Wahl der Parameter für die Ebenenbestimmung getroffen werden, um eine vollständige Erfassung erreichen zu können. Die vorliegenden Ergebnisse deuten darauf hin, dass sich der Ebenenansatz besser für verrauschte und ungleichmäßigen Punktwolken eignet. Der Punktansatz scheint besser für komplexe Räume geeignet. Die Einschränkung der Schnittbedingung für Ebenen und Kanten in beiden Methoden ermöglicht in den verwendeten Datensätzen eine effiziente Vorfilterung, begrenzt dabei jedoch die Allgemeingültigkeit des Modells. Zur zukünftigen Verwendung beider Ansätze sollte insbesondere die Filterung der Ecken und Kanten weiterentwickelt werden. Dazu können die Ansätze z. B. um eine Optimierung der Ecken und Kanten über Graphen erweitert werden. Zusätzlich ist eine Filterung der Kanten mit Hilfe von Punktdichte und -verteilung denkbar. Die Kombination beider Ansätze, kann vermutlich durch die doppelte Beobachtung von Ecken und Kanten die Robustheit und Zuverlässigkeit der Methodik verbessern. Des Weiteren können effizientere Algorithmen zur Bestimmung von Ebenen und Konturpunkten verwendet werden, z. B. das von SOMMER et al. (2020) vorgestellte Local Hough Voting. Dabei sollte der Fokus beim punktbasierten Ansatz besonders auf einer vollständigen Extraktion von Konturpunkten liegen. Da der aktuelle Ansatz nicht automatisch auf die Eingangspunktwolke angepasst wird, ist eine Erweiterung um die Wahl sinnvoller Parameter z. B. auf Basis von Clustering der Flächen und Rauheit der Punktwolke zur Abschätzung der Parameter für RANSAC und MSAC denkbar. Hierzu kann z. B. auch eine semantische Segmentierung zur Ebenendetektion, sowie zur Filterung und Klassifizierung der Objekte im Raum eingesetzt werden (IWASZCZUK et al. 2018). Zur Bestimmung der Beziehungen zwischen detektierten Ecken und Kanten kann ebenfalls Deep Learning verwendet werden. Dabei können die vorgestellten Ansätze dazu dienen, z. B. das von LIU et al. (2021) vorgestellte Verfahren, um die Extraktion verdeckter Ecken zu erweitern. Allerdings müsste dazu ein Trainingsdatensatz aufgebaut werden, der auch verdeckte Kanten enthält.

## 9 Literaturverzeichnis

- AHMED, S. M., TAN, Y. Z., CHEW, C. M., MAMUN, A. A. & WONG, F. S., 2018: Edge and Corner Detection for Unorganized 3D Point Clouds with Application to Robotic Welding. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), <https://arxiv.org/pdf/1809.10468>.
- ARMENI, I., SENER, O., ZAMIR, A. R., JIANG, H.; BRILAKIS, I., FISCHER, M. & SAVARESE, S., 2016: 3D Semantic Parsing of Large-Scale Indoor Spaces. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1534-1543.
- CASTAGNO, J. & ATKINS, E., 2020: PolyLidar3D -- Fast Polygon Extraction from 3D Data. <https://arxiv.org/pdf/2007.12065>.
- COUGHLAN, J. M. & YUILLE, A. L., 1999: Manhattan World: compass direction from a single image by Bayesian inference. Proceedings of the Seventh IEEE International Conference on Computer Vision, (2), 941-947.
- HACKEL, T., WEGNER, J. D. & SCHINDLER, K., 2016: Contour Detection in Unstructured 3D Point Clouds. IEEE Conference on Computer Vision and Pattern Recognition, 1610-1618.



- HIMEUR, C.-E., LEJEMBLE, T., PELLEGRINI, T., PAULIN, M., BARTHE, L. & MELLADO, N., 2022: PCEDNet: A Lightweight Neural Network for Fast and Interactive Edge Detection in 3D Point Clouds. *ACM Trans. Graph.*, **41**(1), 1-21, <https://doi.org/10.1145/3481804>.
- IWASZCZUK, D.; KOCH, T. & STILLA, U., 2017: Innenraumrekonstruktion aus semantisch angereicherten 3D Punkten und Linien. 37. Wissenschaftlich-Technische Jahrestagung der DGPF - Publikationen der DGPF, **26**, 129-137.
- IWASZCZUK, D., KOPPANYI, Z., GARD, N. A., ZHA, B., TOTH, C., & YILMAZ, A., 2018: Semantic Labelling of Structural Elements in Buildings by Fusing RGB and Depth Images in an Encoder-Decoder CNN Framework. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, **42**(1), 225-232, <https://doi.org/10.5194/isprs-archives-XLII-1-225-2018>.
- KHOSHELHAM, K., TRAN, H., DÍAZ-VILARIÑO, L., PETER, M., KANG, Z. & ACHARYA, D., 2018: An Evaluation Framework for Benchmarking Indoor Modelling Methods. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, **42**(4), 297-302, <https://doi.org/10.5194/isprs-archives-XLII-4-297-2018>.
- LÓPEZ IGLESIAS, J., DÍAZ SEVERIANO, J. A., LIZCANO AMOROCHO, P. E., MANCHADO DEL VAL, C.; GÓMEZ-JÁUREGUI, V., FERNÁNDEZ GARCÍA, O., PRECIADOS ROYANO, A. & OTERO GONZÁLEZ, C., 2020: Revision of Automation Methods for Scan to BIM. *International Conference on the Digital Transformation in the Graphic Engineering*, 482-490, [https://link.springer.com/chapter/10.1007/978-3-030-41200-5\\_53](https://link.springer.com/chapter/10.1007/978-3-030-41200-5_53).
- LEHTOLA, V. V., NIKOOHEMAT, S. & NÜCHTER, A., 2021: Indoor 3D: Overview on Scanning and Reconstruction Methods. *Handbook of Big Geospatial Data*, 55-97, [https://link.springer.com/chapter/10.1007/978-3-030-55462-0\\_3](https://link.springer.com/chapter/10.1007/978-3-030-55462-0_3).
- LIU, Y., D'ARONCO, S., SCHINDLER, K. & WEGNER, J. D., 2021: PC2WF: 3D Wireframe Reconstruction from Raw Point Clouds. <https://arxiv.org/pdf/2103.02766>.
- LU, X., LIU, Y. & LI, K., 2019: Fast 3D Line Segment Detection from Unorganized Point Cloud. <https://arxiv.org/pdf/1901.02532>.
- SCHNABEL, R., WAHL, R. & KLEIN, R., 2007: Efficient RANSAC for Point-Cloud Shape Detection. *Computer Graphics Forum*, **26**(2), 214-226. <https://doi.org/10.1111/j.1467-8659.2007.01016.x>.
- SOMMER, C., SUN, Y., GUIBAS, L., CREMERS, D. & BIRDAL, T., 2020: From Planes to Corners: Multi-Purpose Primitive Detection in Unorganized 3D Point Clouds. *IEEE Robot. Autom. Lett.*, **5**(2), 1764-1771, <https://doi.org/10.1109/LRA.2020.2969936>.

# Vergleich verschiedener Unterwasser-Photogrammetrie-Inspektionsmethoden zur Bewertung von unterwasserliegenden Stahlwasserbauwerken

MICHAEL GRÖMER<sup>1</sup> & GOTTFRIED MANDLBURGER<sup>2</sup>

*Zusammenfassung: Die Wartung von Wasserkraftwerken ist mit kostspieligen Inspektionen verbunden. Vor diesem Hintergrund ist die Entwicklung einer neuen Methodik für die Unterwasserinspektion mit Kamerasystemen das Ziel dieser Forschung. Als Studienobjekt wurde ein Unterwasser-Stahlwasserbauwerk eines Laufwasserkraftwerks gewählt. Das Objekt wurde während einer Entleerung mit zwei terrestrischen Laserscannern (RIEGL VZ-400i und Z+F 5010C) im trockenen Zustand vermessen. Zusätzlich wurden an der Sperre Zielmarken angebracht und mit einer Totalstation (Leica TS16) präzise vermessen. Dadurch wurde ein Referenzmodell erstellt, das als exakte geometrische Grundlage für die Beurteilung der in einem zweiten Schritt erstellten photogrammetrischen Modelle dient. Im nächsten Schritt wurde das Objekt im unter Wasser liegenden Zustand aufgenommen. Die Erfassung erfolgte mit einer GoPro Hero Black 7 Kamera von zwei unterschiedlichen Trägerplattformen aus, einer Teleskopstange und einem ferngesteuerten Tauchroboter. Aus diesen Aufnahmen wurden 3D-Modelle erstellt und mit den Referenzdaten verglichen. Es hat sich dabei herausgestellt, dass die Soll-Ist-Abweichungen ohne Verwendung von Passpunkten im Bereich mehrerer Dezimeter liegt und nach deren Einbeziehung in die Bündelblockausgleichung auf wenige cm verringert werden konnte. Die verwendete Trägerplattform hatte einen Effekt auf die erzielte Genauigkeit der 3D-Rekonstruktion, aus den Tauchroboteraufnahmen konnte ein genaueres Modell erstellt werden.*

## 1 Einleitung

Relevante unter Wasser liegende Teile von Wasserkraftwerken werden derzeit in empirisch ermittelten Intervallen einerseits im trockengelegten Zustand und andererseits auch durch Taucher inspiziert. Aufgrund der Tiefe und aus Sicherheitsgründen ist das Tauchen nicht immer möglich oder ratsam. Sowohl Tauchgänge als auch die Entleerung von unterwasserliegenden Anlagenteilen sind zeit- und kostenaufwendig. Außerdem stellen Taucheinsätze immer ein Risiko für die Taucher selbst dar. Unterwasserfahrzeuge, so genannte Remotely Operated Vehicles (ROV), die mit speziellen Inspektionsgeräten ausgestattet sind, könnten in Zukunft einen entscheidenden Beitrag zu effizienteren, sichereren und qualitativ verbesserten Inspektionen in der Wasserkraft leisten (CAPOCCI et al. 2017, GROEMER et al. 2022). Tauchroboter für visuelle Inspektionen oder Vermessungsaufgaben sind im Offshore-Bereich, z.B. in der Gas- und Ölindustrie oder in der Unterwasserarchäologie, bereits etabliert (CHEMISKY et al. 2021). Auf Unterwasserinspektion spezialisierte Unternehmen bieten Trägerplattformen (z.B. ROV, Crawler), Photogrammetrie-Lösungen oder Sonarsysteme und fallweise auch Unterwasser-Lasersysteme an (MAI, C. et al.

---

<sup>1</sup> VERBUND Hydro Power GmbH & TU Wien, Bereich Unterwasserinspektion, E-Werkstraße 31, 8121 Deutschfeistritz, Austria, E-Mail: michael.groemer@verbund.com

<sup>2</sup> TU Wien, Department of Geodesy and Geoinformation, Research Group Photogrammetry, Wiedner Hauptstr. 8-10, 1040 Vienna, Austria, E-Mail: gottfried.mandlburger@geo.tuwien.ac.at

2016). In der Wasserkraft sind solche Technologien noch relativ neu und werden zunehmend für visuelle Inspektionen, aber noch nicht für weitergehende Vermessungsaufgaben eingesetzt. Grund dafür sind unter anderem vor allem die teilweise schwierigen Bedingungen. Vor allem bei Flusskraftwerken sind hier als erschwerende Parameter der hohe Schwebstoffanteil (Trübung) sowie die Strömung zu nennen.

VERBUND arbeitet bereits seit 2018 im Rahmen des Innovations- und Digitalisierungsprogramms "Digitales Wasserkraftwerk" intensiv an der Entwicklung innovativer Methoden zur digital gestützten Unterwasserinspektion. Mit unterschiedlich ausgestatteten ROVs, einer eigens entwickelten Teleskopkamera und einem hochauflösenden Echtzeitsonar stehen mehrere Technologien im Fokus, die im Rahmen der Aktivitäten für die Wasserkraft erprobt und weiterentwickelt wurden und werden. Unter Verwendung verschiedener Trägerplattform/Messsystem-Kombinationen wurden zahlreiche Tests an Wasserkraftwerken durchgeführt. Der Schwerpunkt der F&E-Aktivitäten liegt auf Stereo-Photogrammetrie und auch auf Laservermessung.

Diese neuen Unterwasser-Inspektionsmethoden mit Tauchroboter und Teleskopkamera sind bei VERBUND für verschiedene Routinearbeiten im Bereich der Inspektion von Wasserkraftanlagen mittlerweile etablierte Technik. Für Inspektionsaufgaben mit höherer Anforderung an die Genauigkeit ist es notwendig, die Genauigkeit von 3D Modellen zu ermitteln, die aus Unterwasser-Aufnahmen erstellt wurden. Ein weiterer Fokus der aktuellen Forschung liegt darüber hinaus auf der Reduktion potenzieller Fehlerquellen. In diesem Zusammenhang berichtet der vorliegende Artikel über die Erfahrungen mit Unterwasserinspektion basierend auf Stereophotogrammetrie in der Wasserkraft. Im Mittelpunkt stehen dabei die Fragen, welchen Einfluss (i) die eingesetzten Trägerplattformen (Teleskopstange, ROV) bzw. (ii) Auswertungsansätze (Structure-from-Motion vs. Bündelblockausgleichung mit Passpunkten) auf die Genauigkeit der 3D-Rekonstruktion haben. Der weitere Artikel ist wie folgt gegliedert: In Abschnitt 2 sind das Untersuchungsgebiet, die verwendeten Datensätze inklusive der Referenzmessungen beschrieben. In Abschnitt 3 präsentieren wir die Auswertemethoden. Die erzielten Ergebnisse werden in Abschnitt 4 vorgestellt und kritisch diskutiert. Der Artikel schließt mit einer Zusammenfassung sowie einem Ausblick auf künftige Arbeiten.

## 2 Datensätze und Untersuchungsgebiet

Das Untersuchungsgebiet ist ein Bereich des Oberwasserkanals des Laufwasserkraftwerks Peggau an der Mur (Steiermark, Österreich). Die Position des Kraftwerks ist in Abb. 1 dargestellt. Als spezifisches Studienobjekt wurde die Obertafel des Grundablasses inspiziert, ein unter Wasser liegendes Stahlwasserbauwerk, das in Abb. 3 dargestellt ist.



Abb. 1: Das Laufkraftwerk befindet sich in der Steiermark, nördlich von Graz und ist auf der Abbildung mit einem roten x markiert (N 47°12'; E 15°21'; WGS84). Quelle: <https://basemap.at/>

Die Referenzaufnahmen der Obertafel wurden mit zwei terrestrischen Laserscannern (*RIEGL VZ-400i*, Zoller+Fröhlich 5010C) erstellt (Abb. 2c und Abb. 3). Die Unterwasseraufnahmen wurden mit einer GoPro Hero Black 7 Kamera aufgezeichnet, kombiniert mit diffuser LED Beleuchtung. Die Kamera wurde auf zwei Trägerplattformen montiert. Diese sind in Abb. 2a und b ersichtlich und werden im Abschnitt 3.1 Inspektionsmethoden genauer erläutert.

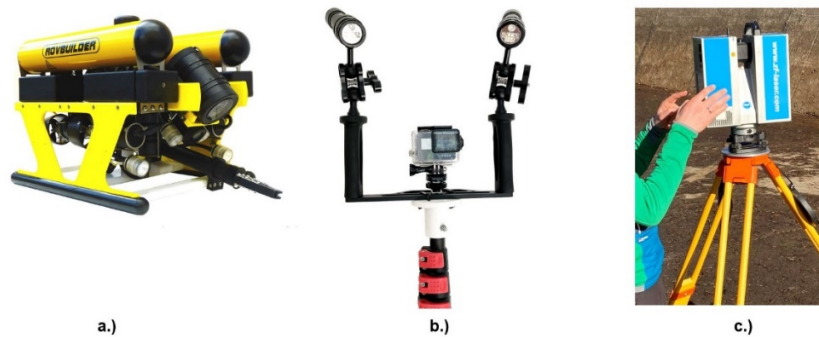


Abb. 2: (a) Tauchroboter, der für die Aufnahme verwendet wurde, (b) GoPro Kamera montiert an der selbst entwickelten Teleskopstange, (c) Zoller + Fröhlich Laserscanner als Basis für das 3D-Referenzmodell.

In Abb. 3 ist der *RIEGL VZ-400i* Laserscanner vor der vermessenen Obertafel abgebildet. Zu erkennen sind auch drei der aufgeklebten Zielmarken, welche mit der Leica TS16 Totalstation eingemessen wurden.



Abb. 3: Laserscan der Obertafel als Referenzmodell, durchgeführt während einer Trockenlegung des Kanals.

## 3 Methoden

### 3.1 Inspektionsmethoden

Bei VERBUND wurden verschiedene Inspektionsmöglichkeiten erprobt und weiterentwickelt, um für die unterschiedlichen Inspektionssituationen eine Vielzahl von Werkzeugen zur Auswahl zu haben. So unterscheidet sich beispielsweise der Einsatz in einem Laufwasserkraftwerk deutlich von dem in einem Pumpspeicherkraftwerk. In diesen beiden Kategorien kann sich die Inspektionssituation vor allem in Bezug auf Trübung, Strömung, Wassertiefe und die zu inspizierenden Anlagenteile unterscheiden. Die zu untersuchenden Objekte haben unterschiedliche Formen und Größen, selbst bei der Inspektion von großen Strukturen wie einem 100 Quadratmeter großen Rechen muss man eine hohe Genauigkeit einhalten und sicherstellen, dass jeder Teil während der Inspektion erfasst wird.

Das Bewegen der Ausrüstung in einem bestimmten Muster, Abstandshalter und spezielle Kameraeinstellungen zusammen mit diffuser Beleuchtung haben sich als hilfreich erwiesen, um die Videoqualität zu verbessern. Die bereits getesteten bzw. verfügbaren Sensoren sind Kameras, Lasersysteme und sonarbasierte Technologie. Je nach Objekt, Genauigkeit, Trübung und Prüftiefe hat jeder Sensor seine Vor- und Nachteile, daher ist die Wahl des richtigen Sensors entscheidend. Um den Sensor zum Objekt zu bringen und ihn in einem bestimmten Muster zu bewegen, wird eine Trägerplattform benötigt. Hier haben sich Tauchroboter, Raupenfahrzeuge, Boote oder eine Teleskopstange als nützlich erwiesen. Mit der Teleskopstange lassen sich Inspektionen besonders gut in Fließgewässern moderater Größe durchführen. ROVs oder Crawler sind mit einer Fernsteuerung ausgerüstet, mit der es möglich ist, bis zu 300 Metern und darüber hinaus zu tauchen.

Mit Hilfe von Unterwasser-Stereophotogrammetrie können aus den aufgenommenen Bildern bzw. Videoframes detaillierte 3D-Modelle erstellt werden. 3D-Modelle verbessern die Inspektionssituation drastisch, da man bei der Betrachtung eines aufgezeichneten Videos von sich

wiederholenden Strukturen leicht die Orientierung verlieren kann. Außerdem muss man aufgrund der Wassertrübung die Kamera oftmals sehr nahe an das zu inspizierende Objekt bringen, was die Auswertung ohne 3D-Modelle schwierig bzw. teilweise unmöglich macht.

Die Tafel wurde mit einer GoPro Hero 7 black Kamera in den folgenden Methoden aufgenommen:

- i. in untergetauchtem Zustand, wobei die Kamera (im Unterwassergehäuse) an einer Teleskopstange befestigt war, und
- ii. in untergetauchtem Zustand mit einem Tauchroboter, an dem die Kamera (in einem Unterwassergehäuse) seitlich montiert wurde.

Der Tauchroboter und die Teleskopkamera sind in Abb. 2 zu sehen.

### 3.2 Auswertungsmethoden

Der Workflow für die Auswertung und den Vergleich zum Referenzmodell ist in Abb. 4 dargestellt. Für die Erstellung der 3D-Modelle wurde die Structure-from-Motion (SfM) Software Agisoft Metashape verwendet (AGISOFT 2022). Der Vergleich zum Referenzmodell erfolgte im Programm CloudCompare (CLOUDCOMPARE 2022). Als Datengrundlage für das Referenzmodell wurde das Lasermodell des Zoller + Fröhlich terrestrischen Laserscanners verwendet.

Bei Unterwasseraufnahmen ist eine hohe Überlappung der Aufnahmen notwendig, da einige erschwerende Faktoren wie das Fehlen von GNSS Daten für die Bildposition und verminderte Bildqualität durch Wassertrübung hinzukommen.

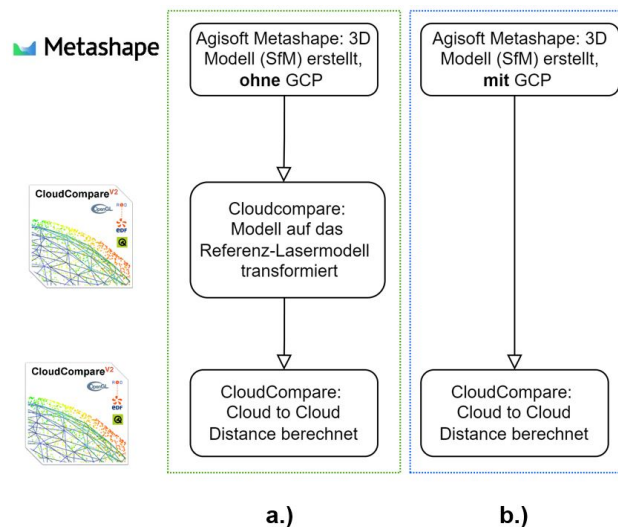


Abb. 4: Die Daten wurden auf zwei Wegen bearbeitet; (a) ohne die Verwendung von Passpunkten (Ground Control Points, GCP) mit anschließender Transformation auf 4 Passmarken; (b) unter Verwendung von Passpunkten in der Bündelblockausgleichung ohne nachfolgende Transformation

Zuerst wurde der in Abb. 4a dargestellte Workflow verwendet, um die 3D-Modelle zu erstellen und zu vergleichen. Die Ergebnisse sind in Abb. 5 und 6 dargestellt und werden in Kapitel 4 diskutiert. Um die Punktwolken für den Vergleich zu referenzieren, wurde die absolute Orientierung des 3D-Modells aus Workflow 4a durch eine Transformation auf vier am Blockrand

befindliche Passpunkte hergestellt. Dabei wurde eine 4x4 Transformationsmatrix verwendet, mit einem variablen Maßstab des Modells.

Beim Workflow, der in Abb. 4b dargestellt ist, ist keine Transformation nötig, da das Modell durch die Passpunkte während der Bündelblockausgleichung in Metashape bereits absolut orientiert wurde. Wie in Kapitel 4 diskutiert wird, sind im Workflow 4a globale, systematische Fehler zu erkennen. Durch die Verwendung von Passpunkten wird versucht, diese während der Auswertung zu reduzieren.

## 4 Ergebnisse und Diskussion

### 4.1 Inspektionsmethoden

Nachdem die 3D-Modelle der zwei verschiedenen Trägerplattformen vorlagen, wurden diese jeweils mit dem Referenz-LasermodeLL verglichen. Es wurde speziell die flächenhafte Modellfehlerverteilung anhand der Cloud-to-Cloud-Distanz (CLOUDCOMPARE 2023) zwischen den einzelnen photogrammetrischen Modellen und dem Laserscanning-Referenzmodell bewertet. Für die beiden photogrammetrischen Blöcke, der mit dem ROV (Abb. 5) und mit der Teleskopstange als Trägerplattform (Abb. 6) erstellt wurden, sind globale Blockverformungen im Dezimeter-Bereich zu beobachten.

Wie in den Abbildungen sichtbar ist, sind die 3D-Modelle kugelförmig konkav gewölbt, wobei dieser Fehler beim 3D-Modell der Teleskopkamera noch stärker auftritt. Bei dem Tauchroboter beträgt der maximale Abstand des 3D-Modells zur Referenz 12 cm, bei der Teleskopkamera 22 cm.

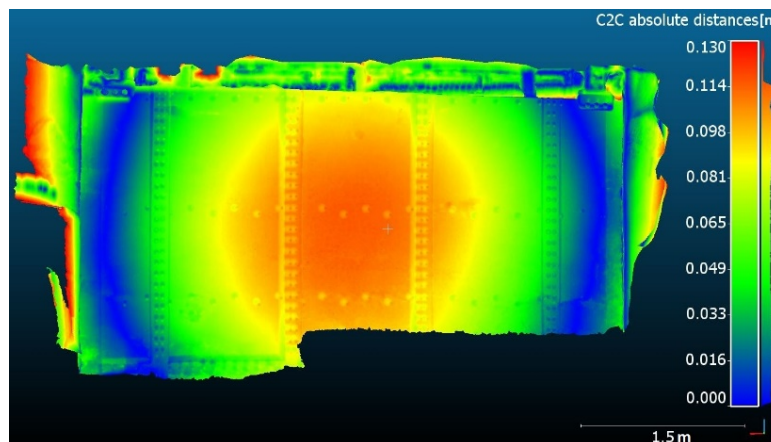


Abb. 5: Vergleich zwischen photogrammetrischen Block (ROV) und Laserscanning-Referenzmodell (ohne Verwendung von Passpunkten).

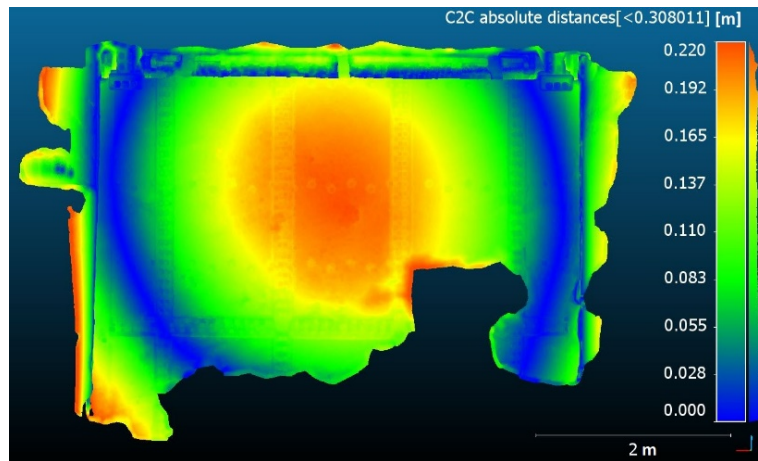


Abb. 6: Vergleich zwischen photogrammetrischem Block (Teleskopkamera) , und dem Laserscanning-Referenzmodell (ohne Verwendung von Passpunkten).

Um diese globalen Blockverformungen zu reduzieren und die konkave Wölbung zu minimieren wurden die mit der Totalstation eingemessenen Passpunkte verwendet und in der Bündelblockausgleichung zur absoluten Orientierung verwendet, um Aufwölbungen zu verhindern. Dafür wurden in einem ersten Schritt sechs gleichmäßig verteilte Passpunkte mit einem zusätzlichen Marker in der Mitte der Platte verwendet. Die Positionen der Marker sind in Abb. 7 dargestellt.

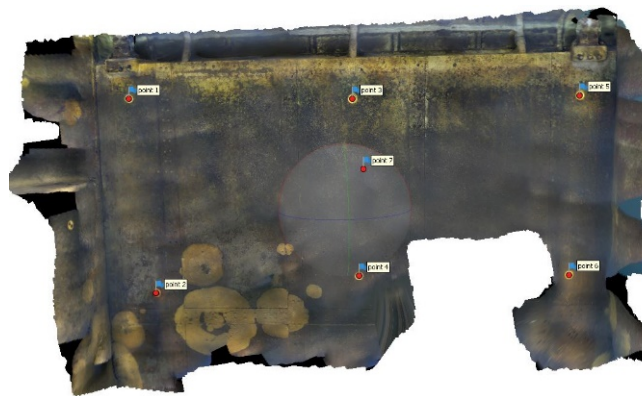


Abb. 7: Positionen der 7 Passpunkte (6x terrestrisch eingemessen + 1x aus Referenzmodell ermittelt) als Basis für absolute Modellorientierung in einer Bündelblockausgleichung.

Die Verwendung von Passpunkten führt zu einer deutlichen Reduktion der globalen Abweichungen. Wie in Abb. 8 abgebildet, sind in diesem Fall die Abweichung zum Referenzmodell im Bereich der Passpunkte weitgehend unter 2 cm. Größere Abweichungen von bis zu 5 cm kommen lediglich am äußersten Modellrand vor.



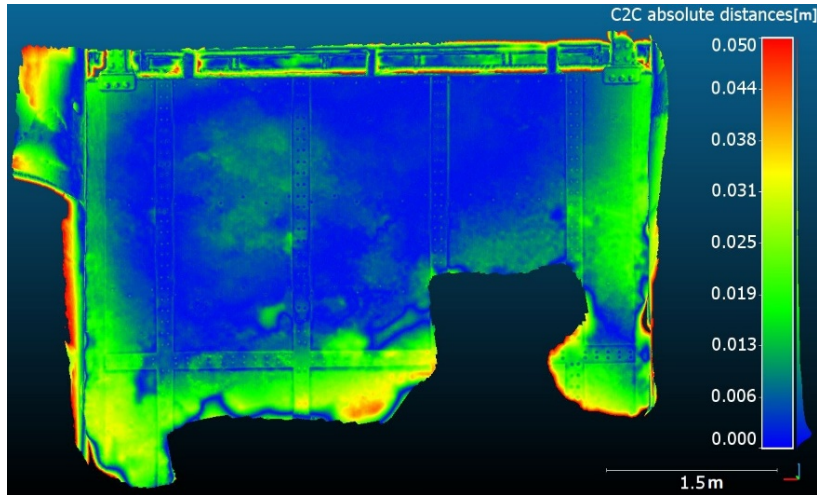


Abb. 8: Vergleich zwischen photogrammetrischem Block (Teleskopkamera) und Laserscanning-Referenzmodell (mit Verwendung von Passpunkten).

Tab. 1: Gegenüberstellung der Abstandparameter der verschiedenen Modelle

	Tauchroboter ohne Passpunkte (Abb. 5)	Teleskopstange ohne Passpunkte (Abb. 6)	Teleskopstange mit Passpunkten (Abb. 8)
Minimaler Abstand	0 m	0 m	0
Maximaler Abstand	0.195 m	0.308 m	0.276 m
Mittlerer Abstand	0.051 m	0.089 m	0.01 m
Standardabw. ( $\sigma$ )	0.067 m	0.083 m	0.035 m

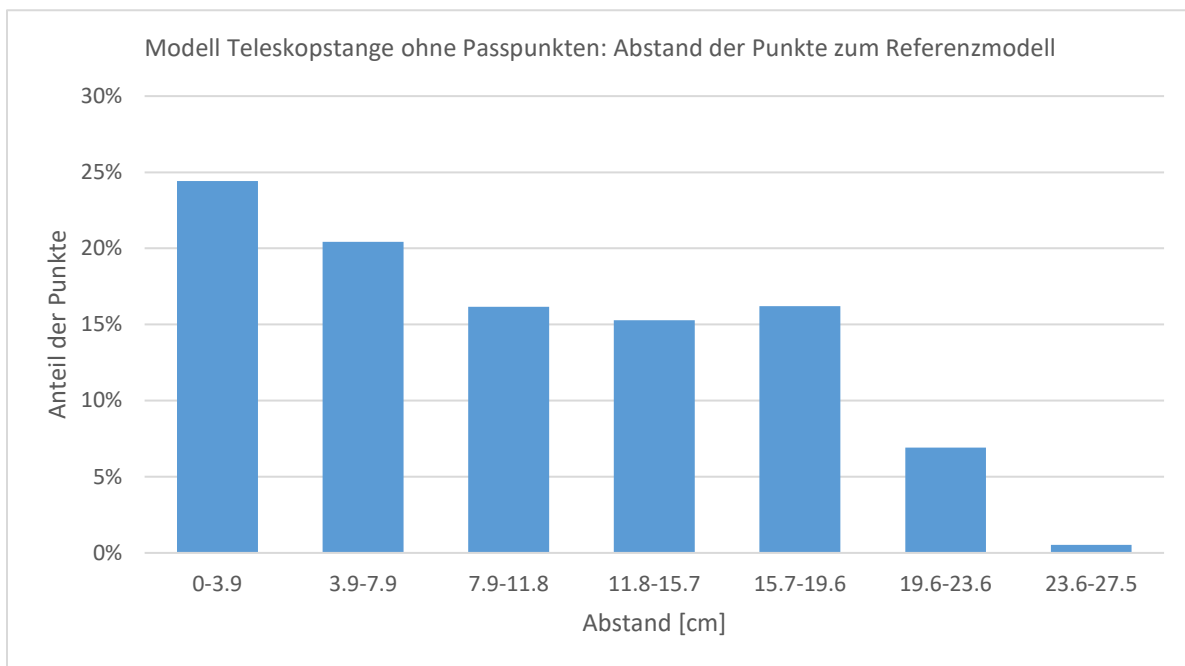


Abb. 9: Histogramm mit den geschätzten Abständen zum Referenzmodell für das 3D-Modell Teleskopstange ohne Passpunkte (Abb. 6)

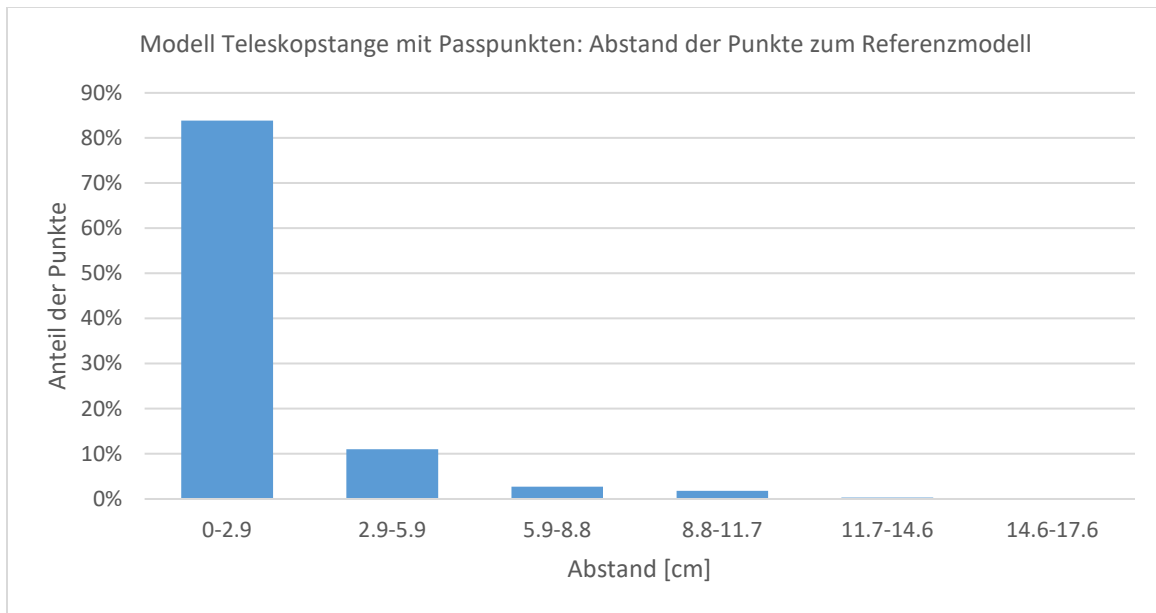


Abb. 10: Histogramm mit den geschätzten Abständen zum Referenzmodell für das 3D Modell Teleskopstange mit Passpunkten (Abb. 8)

## 4.2 Diskussion

Bei der Aufwölbung der Obertafel, wie sie in Abb. 5 und Abb. 6 zu sehen ist, muss es sich um einen Fehler im 3D-Modell handeln. Das kann insofern mit Sicherheit gesagt werden, da der Wasserdruck von dieser Seite auf die Tafel wirkt und zu einer Verformung in die entgegengesetzte Richtung führen würde. Aus statischen Gründen ist hier allerdings nur eine unwesentliche Veränderung anzunehmen. Wenn es lediglich um Visualisierung geht bzw. die Anforderungen an die Genauigkeit und Repräsentationsgüte moderat sind, genügt die Einpassung über vier identische Punkte, wie sie in Abb. 4, Workflow a dargestellt ist. Bei höheren Anforderungen an die geometrische Genauigkeit des rekonstruierten Objekts ist eine Passpunkt-basierte Bündelblockausgleichung unerlässlich (Abb. 4, Workflow b). Aus Tabelle 1 ist ersichtlich, dass sich durch die Hinzunahme von Passpunkten in der photogrammetrischen Auswertung die Fehlermaße deutlich reduzieren. Insbesondere der mittlere Abstand verbessert sich durch die Verwendung von Passpunkten von 8.9 cm auf unter 1 cm. Der maximale Abstand bleibt hoch, da die Ränder des Modells nicht durch Passpunkte korrigiert wurden. Dieses Detail kann in zukünftigen Arbeiten berücksichtigt werden.

In Abb. 9 und in Abb. 10 ist dargestellt, wie groß die Abstände zum Referenzmodell für das Modell Teleskopkamera ohne bzw. mit Passpunkten ist. Hier ist in Abb. 10 klar erkennbar, dass sich der Abstand für den Großteil der Punkte deutlich reduziert hat.

## 5 Zusammenfassung und Ausblick

Bei einem Laufwasserkraftwerk in der Steiermark wurde im entleerten Zustand die Obertafel des Grundablasses vermessen und ein exaktes Referenzmodell erstellt. Dieses wurde verwendet, um die im zweiten Schritt aus Unterwasseraufnahmen erstellten photogrammetrischen 3D-Modelle zu

vergleichen. Dank der rasanten Entwicklung von Photogrammetrie-Software in den letzten Jahren sind innovative Analysen und die Erstellung von 3D-Modellen mit relativ geringem Aufwand und Kosten möglich (REMONDINO et al. 2014). Es hat sich gezeigt, dass die Modelle eine konkave Wölbung aufweisen. Durch die Verwendung von Passpunkten im Zuge der Bündelblockausgleichung konnten die Abweichungen vom Referenzmodell deutlich reduziert und in den cm-Bereich gebracht werden.

Im nächsten Schritt sind weitere Untersuchungen notwendig, wie viele Passpunkte notwendig sind, um Blockaufwölbungen weiter zu reduzieren. Vor allem im Hinblick auf die Untersuchung von größeren unter Wasser liegenden Bauwerken wie bspw. Talsperren ist dieser Punkt von Bedeutung. Zukünftige Untersuchungen zielen darauf ab, eine Prüfmethode zu entwickeln, die die Prüfung von getauchten Stahlwasserbaukomponenten mit einer spezifizierten Genauigkeit von < 1 cm ermöglicht.

## 6 Literaturverzeichnis

- AGISOFT METASHAPE PROFESSIONAL (Version 1.8.3) (Software) (2022). Retrieved from <http://www.agisoft.com/downloads/installer/>
- CAPOCCI, R., DOOLY, G., OMERDIC, E., COLEMAN, J., NEWE, T. & TOAL, D., 2017: Inspection-class remotely operated vehicles—a review. *Journal of Marine Science and Engineering*, **5**(1), <https://www.mdpi.com/2077-1312/5/1/13>.
- CHEMISKY, B., MENNA, F., NOCERINO, E. & DRAP, P., 2021: Underwater Survey for Oil and Gas Industry: A Review of Close Range Optical Methods. *Remote Sensing*, **13**(14). <https://www.mdpi.com/2072-4292/13/14/2789>.
- CLOUDCOMPARE (version 2.12) (GPL software) (2022). Retrieved from <http://www.cloudcompare.org/>
- GROEMER, M. & ARTMANN, M., 2022: Digitally assisted underwater inspections for hydropower. *Hydropower & Dams*, **29**(5).
- MAI, C., PEDERSEN, S., HANSEN, L., JEPSEN, K. L. & YANG, Z., 2016: Subsea infrastructure inspection: A review study. *IEEE International Conference on Underwater System Technology: Theory and Applications (USYS)*, 71-76.
- REMONDINO, F., SPERA, M. G., NOCERINO, E., MENNA, F. & NEX, F., 2014: State of the art in high density image matching. *The Photogrammetric Record*, **29**(146), 144-166.

# UWSensor: Ein optisches Unterwasser-3D-Scannersystem für mobile Kartierung

**MICHAEL BLEIER<sup>1</sup>, CHRISTOPH MUNKELT<sup>2</sup>, MATTHIAS HEINZE<sup>2</sup>,  
CHRISTIAN BRÄUER-BURCHARDT<sup>2</sup>, STEFAN MARX<sup>3</sup>, DIRK FROMME<sup>4</sup>, CHRISTIAN DETTMER<sup>5</sup>,  
HELGE ANDREAS LAUTERBACH<sup>1</sup> & ANDREAS NÜCHTER<sup>1</sup>**

*Zusammenfassung: Im Projekt UWSensor wurde ein neuartiger optischer 3D-Unterwassersensor basierend auf Streifenlichtprojektion entwickelt. In den Unterwassergehäusen des Sensors sind eine Stereokamera und ein GOBO-Projektor für die 3D-Messung, sowie eine Farbkamera und eine inertielle Messeinheit (IMU) integriert. Das Sensorsystem ermöglicht damit eine flächige 3D-Erfassung in einem Messvolumen von ca. 1m × 1m × 1m, sowie eine gleichzeitige Erfassung von Farbdaten. Die Musterprojektion ist vom Farbspektrum und durch den Einsatz von Hochleistungs-LED-Einheiten an den Unterwasserkontext angepasst. Die Vorteile des Sensorsystems sind eine sehr hohe Genauigkeit und eine 3D-Abtastrate von bis zu 60Hz. Die flächige Antastung bei gleichzeitig hoher Update-Rate ermöglicht eine fortlaufende Registrierung von Einzelscans. Dies ermöglicht den Einsatz zur mobilen Kartierung und 3D-Erfassung. Dieser Beitrag beschreibt das Gesamtsystem und die Datenverarbeitung, sowie erste Ergebnisse aus Offshore-Tests des Sensorsystems mit einem ferngesteuerten Unterwasserfahrzeug in der Ostsee.*

## 1 Einleitung

Das Interesse an präziser und zuverlässiger Unterwassermesstechnik ist groß. Beispielsweise wollen Archäologen Artefakte in situ mit hoher Detailtreue dokumentieren und in der Meeresforschung benötigen Biologen Messwerkzeuge zur Beobachtung des Korallenwachstums.

Auch Geologen sind auf Messdaten für die Modellierung des Sedimenttransports angewiesen. Darüber hinaus ist für die Errichtung von Offshore-Bauwerken, sowie deren Wartung und Überwachung eine millimetergenaue Vermessung von vorhandenen Strukturen und Defekten unerlässlich. Hochauflösende 3D-Scanner ermöglichen eine kosteneffektive Maßkontrolle und Inspektion von Strukturen im Wasser, beispielsweise auf Lochfraßkorrosion.

Darüber hinaus müssen im Meer installierte Bauteile vorher passgenau an Land gefertigt werden, da eine Anpassung im Wasser nur schwer möglich ist. Bei Pipeline Verbindungs- und Anschlussstücken ist hierbei präzise Messtechnik nötig, um eine leckdichte Verbindung zu gewährleisten

---

<sup>1</sup> Informatik XVII Robotik, Julius-Maximilians-Universität Würzburg, Am Hubland, D-97074 Würzburg, E-Mail: [michael.bleier, helge.lauterbach, andreas.nuechter]@uni-wuerzburg.de

<sup>2</sup> Fraunhofer-Institut für Angewandte Optik und Feinmechanik IOF, Albert-Einstein-Str. 7, D-07745 Jena, E-Mail: [Christoph.Munkelt, Matthias.Heinze, Christian.Braeuer-Burchardt]@iof.fraunhofer.de

<sup>3</sup> Oktopus GmbH, Wischhofstr. 1-3 G.13, D-24148 Kiel, E-Mail: smarx@oktopus-kiel.eu

<sup>4</sup> 3plusplus GmbH, Xantener Straße 191, D-47665 Sonsbeck, E-Mail: Dirk.Fromme@3plusplus.com

<sup>5</sup> SeaReenergy Offshore Holding GmbH & Cie. KG, Alsterufer 26, D-20354 Hamburg, E-Mail: C.Dettmer@SeaReenergy.com

trotz der Ausdehnung aufgrund von Druck und Temperatur der transportierten Flüssigkeiten oder Gase. Die Genauigkeitsanforderungen sind hoch.



Abb. 1: UWSensor montiert im hydraulischen Manipulator eines Tauchroboters von Baltic Taucherei- und Bergungsbetrieb Rostock.

Beispielsweise werden für die Vermessung der Lage von Anschlussstellen von Pipelines eine relative Genauigkeit von wenigen Zentimetern und Winkelfehler von weniger als  $0,5^\circ$  pro 100m Distanz gefordert. Dies wird in der Industrie unter anderem mit markerbasierter Nahbereichspho- togrammetrie erreicht. Allerdings benötigt dies zusätzliche Infrastruktur in Form von optischen Markern, welche vorher am Meeresboden ausgebracht werden müssen.

Im Rahmen des UWSensor-Projekts wurde ein neuer optischer 3D-Unterwassersensor für die mobile Offshore-Vermessung entwickelt. Der Sensor ist derart konzipiert, dass er für Messungen unter Einsatz eines ferngesteuerten Unterwasserfahrzeugs konfiguriert werden kann. Abbildung 1 zeigt den Sensor montiert im hydraulischen Manipulatorarm eines Tauchroboters. Der Sensor arbeitet hierbei in Bewegung und die Einzelscans werden miteinander registriert, um das Scanfeld kontinuierlich zu erweitern. Das Ziel ist eine 3D-Erfassung mit Bewegungsgeschwindigkeiten von bis zu 1 Meter pro Sekunde.

Die Vorteile des neu entwickelten Unterwasser-3D-Sensorsystems sind:

- Flächige Abtastung in einem Messvolumen von ca.  $1 \times 1 \times 1$  m bei einem typischen Messab- stand im Bereich 0,5 – 2,5 m ermöglicht die Registrierung von Einzelscans
- Sehr hohe Genauigkeit mit einem systematischem Messfehler kleiner als 0,2 mm für stati- sche Scans in klarem Wasser mit einem Messabstand von bis zu 2,4 m (BRÄUER-BURCH- ARDT et al. 2022a)
- Hohe Punktdichte (ca. 1 Mio. Punkte pro Scan) und eine 3D-Abtastrate von bis zu 60 Hz
- Simultane Aufnahme von 3D-Punktwolken und hochaufgelösten Farbbildern mit LED-Blitz
- Automatische Bewegungskompensation basierend auf visueller und inertialer Odometrie während der Datenaufnahme
- Benutzerinterface mit Echtzeitvisualisierung zur Qualitätskontrolle und Unterstützung des Fernbedieners des Tauchroboters während der Datenaufnahme

## 2 Unterwasser 3D-Sensor

Der im Forschungsprojekt UWSensor neu entwickelte 3D-Sensor basiert auf aktivem Stereo mit einer aperiodischen Streifenlichtprojektion (HEIST et al. 2018). Hierzu wird ein Graphical Optical Blackout (GOBO) Projektor mit einer leistungsfähigen LED eingesetzt, um ausreichend Musterkontrast bei hoher Bildwiederholrate im Wasser zu erreichen. Die 3D-Daten Generierung basiert auf einem Mehrbildverfahren, bei dem typischerweise aus einer Sequenz von 10 Stereobildern eine Punktwolke rekonstruiert wird. Der Sensor erfasst kontinuierlich 3D-Daten in einem Bereich von ca. 1 m<sup>2</sup> bei einer Messdistanz von 2 m. Zwischen den einzelnen Stereobildsequenzen werden mit einer konstanten Datenrate von 25 Hz Farbbilder mit 7 MPixel aufgenommen. Diese decken einen größeren Sichtbereich von etwa 2×2 m bei einer Entfernung von 2 m ab. Hierzu werden die Kameras von einer zentralen Steuereinheit in einem alternierenden Betrieb getriggert. Zeitlich synchron mit den Farbbildern wird ein LED-Blitz ausgelöst, respektive die LED-Beleuchtung des Musterprojektors während der Stereoaufnahmen aktiviert. Zwischen den einzelnen Belichtungen sind kleine Totzeiten konfiguriert, um einen Lichteintrag durch Nachleuchten des LED-Blitz in die Musterprojektion zu minimieren.

### 2.1 Aufbau des Sensorsystems

Die Sensorik ist in vier Unterwassergehäusen integriert, die in Abbildung 2 zu sehen sind. In den äußeren beiden Gehäusen sind zwei monochrome Messkameras als Stereopaar verbaut. Diese werden in Kombination mit in einem der zentralen Gehäuse verbauten Musterprojektor für die 3D-Messung eingesetzt. Der Projektor ist eine Eigenentwicklung von Fraunhofer IOF und besteht aus einer im blauen Wellenlängenbereich arbeitenden LED mit hoher Lichtleistung, einer rotierenden Glasmaske (GOBO Rad), Motor und Projektionslinse. Zusätzlich sind im gleichen Gehäuse die elektrischen Treiber und Steuerelektronik integriert. Neben dem Unterwassergehäuse für den Projektor befindet sich ein weiteres Gehäuse, das die Farbkamera und eine inertielle Messeinheit (IMU) mit faseroptischen Gyroskop enthält. Die Unterwassergehäuse sind gemeinsam auf einer Carbonstange montiert, welche vom Trägerrahmen entkoppelt ist. Hierdurch wird eine gute Kalibrierstabilität des Sensorsystems erreicht werden. Zusätzlich befinden sich seitlich außen am Sensorträger zwei LED-Blitzlichter. Diese werden so angeordnet, dass sie primär das Messobjekt ausleuchten und nicht Partikel im Nahbereich der Kamera beleuchten, um Rückstreuung zu minimieren. Am Sensorträger befindet sich ein weiteres Unterwassergehäuse mit der Elektronik für die Spannungsversorgung und Netzwerktechnik für die Datenübertragung.

Sämtliche Sensordaten werden über Netzwerk übertragen. Hierzu wird beim Messeinsatz eine Glasfaser der Tether des Tauchroboters mitbenutzt. Die 3D-Berechnung und Auswertung der Daten erfolgt außerhalb des Wassers auf einer PC-Workstation am Schiff, die über die 10-Gigabit Glasfaserstrecke direkt mit dem Sensor verbunden ist. Hierdurch kann darauf verzichtet werden leistungsfähige Hardware für die Berechnung in den Unterwassergehäusen zu integrieren wodurch Kosten sowie Gewicht eingespart werden. Die 3D-Rekonstruktion und Teile der Weiterverarbeitung erfolgen dabei online, um eine direkte Qualitätskontrolle der Aufnahmen zu ermöglichen.



Abb. 2: UWSensor 3D-Sensorsystem in einem Wassertank bei Versuchen zur Kalibrierung.

## 2.2 Benutzerinterface

Ein wichtiger Punkt für den effektiven Einsatz des Sensors in einer Messkampagne ist das Benutzerinterface. Hierzu wurde im Projekt eine Bediensoftware von 3plusplus entwickelt, welche sowohl eine Visualisierung für den Piloten des Unterwasserfahrzeugs als auch die Schnittstelle für den Bediener des Sensors bereitstellt. Ein Beispiel für eine Ansicht der Nutzersoftware ist in Abbildung 3 dargestellt.

Die Software dient der Parametrisierung des Sensors, um schnell Belichtungszeit und Lichtempfindlichkeit an die Gegebenheiten im Wasser anzupassen. Hierbei unterstützen Visualisierungen, die beispielsweise über- bzw. unterbelichtete Regionen im Bild markieren sowie eine direkte Anzeige des 3D-Rekonstruktionsergebnisses für die Qualitätskontrolle bereitstellen. Des Weiteren erlaubt das Bedienprogramm eine Steuerung der Aufnahme des Datenstroms sowie das Auslösen von Einzelaufnahmen, z.B. für die Kalibrierung des Sensors.

Für den Fernbediener des Tauchroboters wird eine latenzarme Darstellung des Farbbilddatenstroms bereitgestellt, die dieser zur Navigation mitbenutzen kann. Um den Piloten zusätzlich zu unterstützen einen möglichst optimalen Messabstand zum Objekt einzuhalten, werden verschiedene optische Hilfen dargestellt, die schnell erfassbar und interpretierbar sind. Für den Piloten wird auf eine Visualisierung der 3D-Daten verzichtet, da diese einen kleineren Sichtbereich haben und schwieriger interpretierbar sind als die Farbbilder.

Für die Visualisierung des registrierten Gesamtmodells wurde eine browser-basierte Lösung unter der Nutzung von Potree (SCHÜTZ 2016) entwickelt. Dies beinhaltet eine Nutzerverwaltung, Organisation in Scanprojekte und automatische Erstellung von Voransichten der Punktwolken. Hierbei wurde gezielt auf Webtechnologie gesetzt, um einen einfachen Fernzugriff und eine Betrachtung der Daten ohne spezielle 3D-Software zu ermöglichen. Dies erlaubt zum einen, dass alle Beteiligten und Interessenten unabhängig von ihrer Vorerfahrung die Daten betrachten und untersuchen können. Zum anderen erlaubt es perspektivisch eine Übertragung der Daten über Mobilfunk oder Satellitenlink vom Messort ins Büro. Hierdurch können die Daten parallel ausgewertet werden und der Personaleinsatz auf dem Schiff kann dadurch während der Messkampagne potenziell verringert werden.

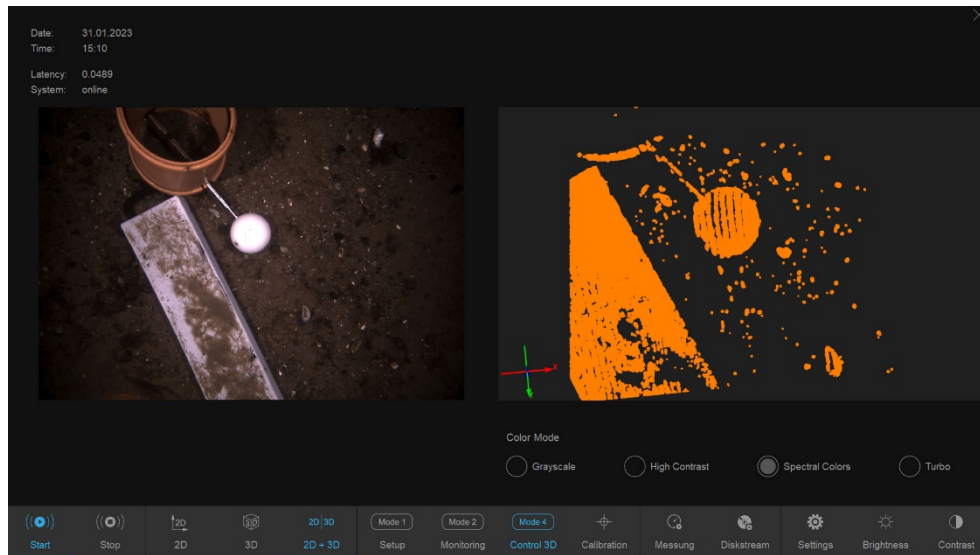


Abb. 3: Benutzerinterface des Sensors. Links: Farbbild des Sensors. Rechts: 3D-Punktwolke.

### 3 Datenverarbeitung für mobile Kartierung

Das System realisiert eine Bewegungsschätzung sowie eine Fusion der Einzelscans in ein global konsistentes 3D-Modell. Hierbei wird zunächst visuelle Odometrie in Kombination mit einem faseroptischen Gyroskop genutzt, um die Eigenbewegung des Sensors zu schätzen. Die daraus latenzarm abgeleiteten Posen werden dem Live-Aufnahmesystem des Sensors bereitgestellt, um eine Bewegungskompensation der 3D-Scans in Echtzeit zu realisieren. Anschließend wird eine mehrstufige Registrierungsstrategie angewendet, um die Schätzung der Trajektorie weiter zu verfeinern. Anhand der bestimmten Sensortrajektorie werden die Einzelscans zu einem Gesamtmodell aggregiert und in reduzierter Auflösung visualisiert, wodurch eine Qualitätskontrolle während der Datenaufnahme ermöglicht wird.

#### 3.1 3D-Berechnung und Kalibrierung

Die 3D-Messung wird anhand des Stereobilddatenstroms mit Aufnahmen des aperiodischen Streifenmusters des GOBO-Projektors durchgeführt. Die 3D-Daten werden hierbei durch Stereotriangulation anhand der kalibrierten Messkameras berechnet. Die Korrespondenzsuche zwischen den Stereobildsequenzen basiert dabei auf der höchsten Korrelation des zeitlich veränderlichen Grauwertverlaufs der Pixel entlang der Epipolarlinie (HEIST et al. 2018).

Für den Unterwassereinsatz wird dabei ein erweitertes Lochkameramodell eingesetzt, das ausgehend von der geometrischen Modellierung des Strahlverlaufs die Brechung an den Medienübergängen zwischen Luft, Glas und Wasser berücksichtigt (BRÄUER-BURCHARDT et al. 2022b). Längenabweichungen in Abhängigkeit der Messdistanz werden zusätzlich durch eine Fehlerausgleichsfunktion im 3D-Objektraum kompensiert (BRÄUER-BURCHARDT et al. 2022a).

Die Kalibrierung der Stereokameras wird dabei anhand eines Kreismusters durchgeführt. Im Wasser wird hierzu die Kalibrierstruktur auf den Grund gelegt und mit dem Sensor ca. 10 – 20 Aufnahmen aus verschiedenen Winkeln erstellt. Hierbei kommen zusätzliche ArUco Marker auf dem



Kalibriermuster zum Einsatz, um jedem Kreis eine eindeutige ID zu geben. Die Kalibrierung wird mit BINGO (KRUCK 2014) als Bündelblockausgleichssoftware durchgeführt.

Die Stereokalibrierung wird sowohl zwischen den beiden monochromen Messkameras, als auch zwischen der linken Messkamera, die das Bezugskordinatensystem für die 3D-Messung darstellt, und der Farbkamera berechnet. Hierdurch werden im gleichen Durchgang die Parameter für das Kameramodell der Farbkamera als auch die Transformation zwischen dem Koordinatensystem von Farb- und 3D-Aufnahme bestimmt.

### **3.2 Visuelle und inertielle Odometrie**

Zur Erstellung einer latenzarmen Trajektorienschätzung des Sensors unabhängig von den 3D-Daten wird visuelle und inertielle Odometrie (VIO) eingesetzt. Das Verfahren basiert hierbei auf VINS-Mono (QIN et al. 2018). VINS-Mono verwendet den Harris-Feature-Detektor (auch als Shi-Tomasi Methode bekannt), für die automatische Extraktion von Bildmerkmalen für visuelle Odometrie.

Zu Beginn der Aufnahme muss die Posenschätzung initialisiert werden. Dies bedeutet, dass die Methode nicht direkt aus dem Stillstand verwendet werden kann. Der Sensor muss zunächst bewegt werden, um genug Parallaxe zu erzeugen, dass eine initiale, unskalierte Structure from Motion (SfM) Rekonstruktion der Merkmalspunkte möglich ist. Diese wird dann durch eine Anpassung mit den aufintegrierten IMU-Daten skaliert. Sobald dieser Initialisierungsprozess abgeschlossen ist, kann mit der eigentlichen Messaufnahme begonnen werden. Dies erfolgt typischerweise nach ca. 2-4 Sekunden abhängig von der Bewegungsgeschwindigkeit des Sensors.

Die Kalibrierung der relativen Lage und Orientierung zwischen Kamera und IMU wird in Luft bestimmt. Im Wasser bleiben die Parameter der Kamera-IMU-Transformation fixiert. Bei Aufnahmen im Wasser werden typischerweise aufgrund der Kinematik von Unterwasserfahrzeugen keine Rotationen um alle Sensorachsen erzielt. Dies führt zu einer ungenauen Schätzung insbesondere der Translation zwischen Kamera und IMU. Des Weiteren ist eine schnelle Initialisierung der visuellen Odometrie auch bei nahezu linearen Trajektorien möglich, wenn die Kamera-IMU-Transformation als bekannt vorausgesetzt wird. Die stochastischen Fehlerparameter wurden anhand des Funktionsgraphen der Allan-Varianz (HOU & EL-SHEIMY 2003) einer Aufnahme des IMU-Datenstroms über mehrere Stunden bestimmt. Diese dienen hier aber primär als Gewichte für die Abwägung zwischen Kamera und IMU-Messung und nicht als vollständiges Modell des Rauschverhaltens und der Messunsicherheit.

Die Zeitverschiebung zwischen Kamera und IMU-Datenstrom wird online bestimmt, da diese hardwaretechnisch keine präzise gemeinsame Zeitbasis besitzen. Die relative Bewegung wird dann durch Ausgleich der visuellen und IMU Messungen in einer lokalen Umgebung bestimmt. Die Optimierung wird hier lokal über eine fortlaufende Gruppe von Keyframes durchgeführt. Keyframes werden selektiert anhand eines Schwellwerts auf die mittlere Parallaxe der einzelnen Merkmalspunkte.

Zu beachten ist, dass in diesem Schritt keine Relokalisierung oder ein Schließen von Schleifen anhand visueller Merkmale eingesetzt wird. Das Ziel ist eine stetige Trajektorie ohne Sprünge zu

erhalten, da die visuelle Odometrie primär zur lokalen Bewegungskompensation eingesetzt werden soll. Ein eventuell durch Restfehler entstehender Drift wird später in der 3D-Registrierung kompensiert.

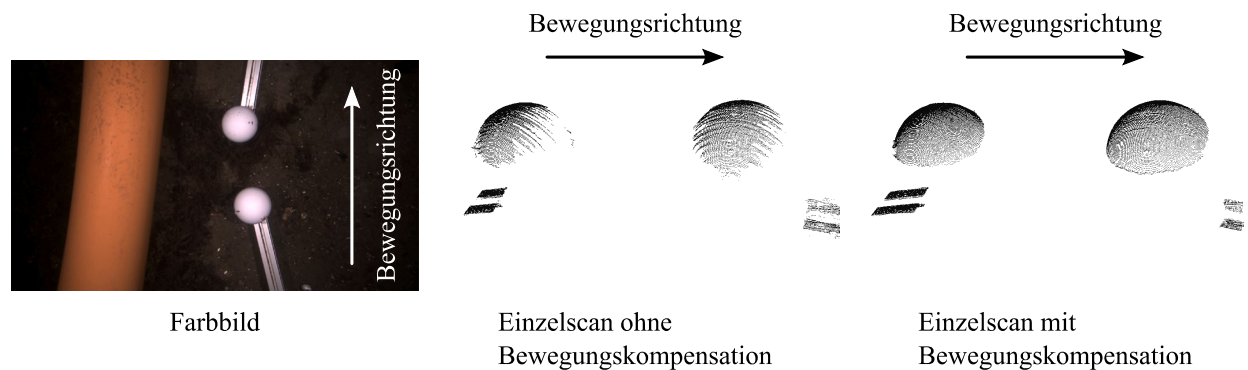


Abb. 4: Effekt der Bewegungskompensation auf das Ergebnis eines Einzelscans. Links: Farbbild der Messszene. Mitte: Scan ohne Berücksichtigung der Bewegung. Rechts: Verbessertes und vollständigeres Ergebnis mit höherer Punktdichte durch Bewegungskompensation.

### 3.3 Bewegungskompensation der 3D-Scans

Wie in Abschnitt 3.1 beschrieben nutzt die 3D-Berechnung eine zeitliche Korrelation über einen Stapel von Grauwertbildern von linker und rechter Messkamera. Der Ansatz setzt hierbei voraus, dass die Messszene statisch ist. Diese Annahme kann beim Einsatz für die mobile Kartierung nicht erfüllt werden. Eine quasi-statische Messsituation durch Hochgeschwindigkeits-Projektion und Aufnahmen ist im Wasser nicht erzielbar. Durch die Absorption des Lichts im Wasser, Trübung und begrenzte Leistung der Projektionseinheit sind längere Integrationszeiten der Kameras notwendig.

Deshalb wird im UWSensor der Ansatz verfolgt die Bewegung algorithmisch zu kompensieren. Der Ansatz für die Bewegungskompensation ist in (BRÄUER-BURCHARDT et al. 2022a) und (BLEIER et al. 2022) beschrieben. Hierbei wird für die Zeitspanne der Aufnahme der Messbilder eine näherungsweise lineare Bewegung des Sensors angenommen. Anhand der Bewegungsschätzung der visuellen Odometrie und einer angenommenen mittleren Messentfernung wird dann die Verschiebung im Bildraum approximiert. Um den berechneten Pixel-Offset werden dann die Grauwertbilder korrigiert, bevor die Korrespondenzsuche angewandt wird.

Dies führt zu einer deutlichen Verbesserung des Messergebnisses in Bewegung. Sowohl in Bezug auf die Vollständigkeit der Punktwolke, da mehr Punkte rekonstruiert werden können, als auch die Genauigkeit der Messungen. Abbildung 4 zeigt ein Beispiel für eine Überfluggeschwindigkeit des Sensors von ca. 0,25 m/s. Das linke Bild zeigt das Farbbild der Messszene mit zwei Kugeln. In der Mitte ist das 3D-Rekonstruktionsergebnis ohne Berücksichtigung der Bewegung dargestellt. Die Kugeloberflächen können nicht vollständig erfasst werden und weisen streifenförmige Lücken auf. Im rechts dargestellten Ergebnis mit Bewegungskompensation sind die Kugeloberflächen mit deutlich höherer Punktdichte und gleichmäßig ohne Lücken erfasst.

### 3.4 Registrierung der Punktwolken

Aufbauend auf der initialen Schätzung der Trajektorie mittels visueller Odometrie wird eine mehrstufige Registrierungsstrategie angewandt, um die Trajektorie weiter zu verfeinern. Abbildung 5 gibt einen Überblick der Verarbeitungsschritte. Die ersten beiden Stufen werden zeitgleich zur Datenaufnahme ausgeführt und erlauben dem Operator eine erste Einschätzung der Ergebnisse. In der Nachverarbeitung erfolgt eine globale Optimierung der 3D-Punktwolke.

Zur Vorbereitung für die Registrierung, werden die 3D Daten zunächst gefiltert, um die Größe der zu verarbeitenden Punktwolke zu reduzieren und dabei eine ausgewogene räumliche Verteilung der gewählten Punkte zu gewährleisten. Mittels des Iterative Closest Point (ICP) Algorithmus wird jede 3D Punktwolke sequenziell gegen ihre Vorgänger registriert. Ziel dieses Schrittes ist es, die Trajektorie lokal zu optimieren und damit die resultierende 3D Karte zu verbessern. Der Suchradius wird dabei in Abhängigkeit der Geschwindigkeit des Sensors sehr klein gewählt (im Bereich weniger cm) um der Gefahr von groben Registrierungsfehlern aufgrund des beschränkten Sichtfeldes sowie der zu erwartenden Strukturarmut der erfassten Szenerie zu begegnen.

Die Registrierung der Einzelscans ist implizit davon abhängig, dass ausreichend 3D-Struktur in der Punktwolke vorhanden ist, damit die ICP-Registrierung konvergiert. Abhängig von der erfassten Messszene ist dies nicht immer gegeben. Wird beispielsweise nur die zylinderförmige Oberfläche eines Rohres erfasst, dann ist die Registrierung in Richtung des Rohres nicht beschränkt. In diesen Fällen kann die ICP-Registrierung deaktiviert werden und es wird in diesem initialen Schritt ausschließlich die Trajektorien-schätzung der visuellen Odometrie verwendet.

Da sich verbleibende Restfehler in der Registrierung akkumulieren und der bestehende Drift der visuellen Odometrie nicht vollständig eliminiert wird, erfolgt im nächsten Schritt eine zweite Registrierung mit einem zeitkontinuierlichen ICP-Verfahren. Grundgedanke ist, dass der Fehler der Trajektorie in zeitlicher Nähe einer betrachteten Pose vernachlässigbar ist. Die Trajektorie wird dann in Teilabschnitte zerlegt und mehrere aufeinander folgende 3D Scans um einen gewählten Referenzscan zu einer Teilkarte zusammengefasst. Die Teilkarten werden wiederum gegen ihre Vorgänger registriert. Die Änderung der Pose eines Referenzscans wird im Anschluss auf die Posen zwischen zwei Referenzscans verteilt, um die Kontinuität der Trajektorie zu wahren. Für kleine Änderungen ist dabei eine lineare Verteilung der Translation bzw. eine Verteilung der Rotationsfehler mit Spherical Linear Interpolation (SLERP) ausreichend. Zur Korrektur des akkumulierten Drifts werden Schleifen erkannt und geschlossen (SPRICKERHOF et al 2011). Dazu werden die Posen der aggregierten Teilkarten optimiert und die Änderungen anschließend analog zu oben auf die Einzelposen verteilt.

Die resultierende Karte dient der Visualisierung während des Messvorgangs. Um die Datenmenge für die Visualisierung zu reduzieren, werden die Einzelscans ausgedünnt. Hierzu wird zum Gesamtmodell jeweils nur dann ein neuer Scan hinzugefügt nachdem sich das Sensorsystem mindestens 5-10 cm bewegt hat, je nach gewünschter Punktdichte. Die aus den aggregierten Einzelscans anhand der bestimmten Trajektorie zusammengesetzte 3D-Karte wird nachfolgend in einem Octree-basierten Verfahren in der räumlichen Auflösung reduziert und der Visualisierung zur Verfügung gestellt. Dies passiert asynchron zur Trajektorien-schätzung und Registrierung. Der Visualisierung wird anschließend ein aktualisiertes Gesamtmodell alle ca. 1-5 Sekunden zur Verfügung gestellt.

Die Nachverarbeitung der Daten zur Erstellung der finalen 3D Punktwolke erfolgt mit zeitkontinuierlichem SLAM (ELSEBERG et al. 2013). Die überlappenden Teilkarten werden in diesem Schritt mit Graph basiertem SLAM rigide eingepasst. Das Resultat dient als Startwert für die semi-rigide Registrierung zur numerischen Minimierung des Trajektoriefehlers.

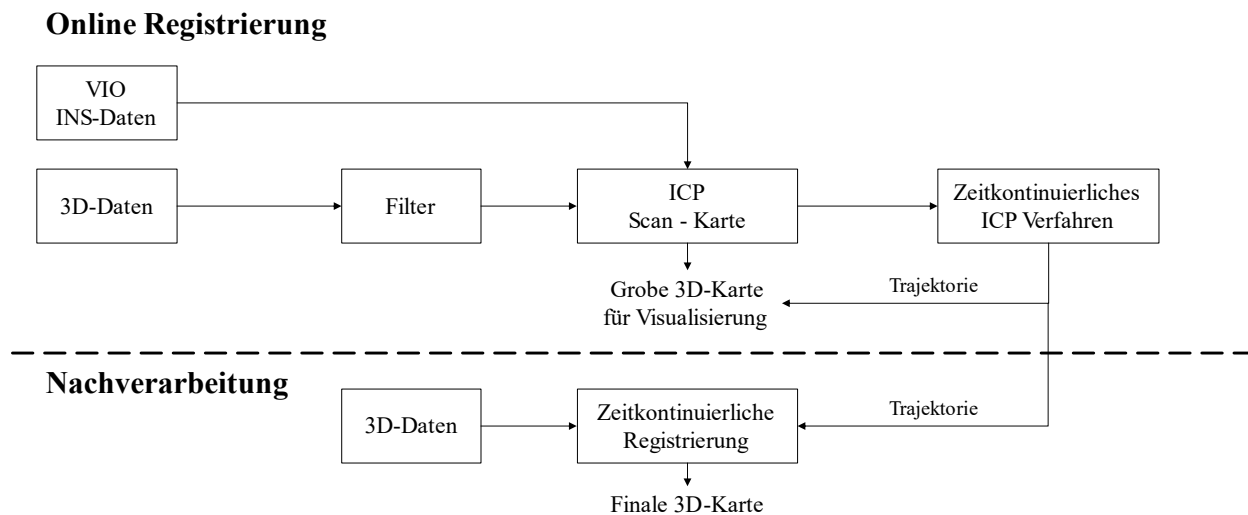


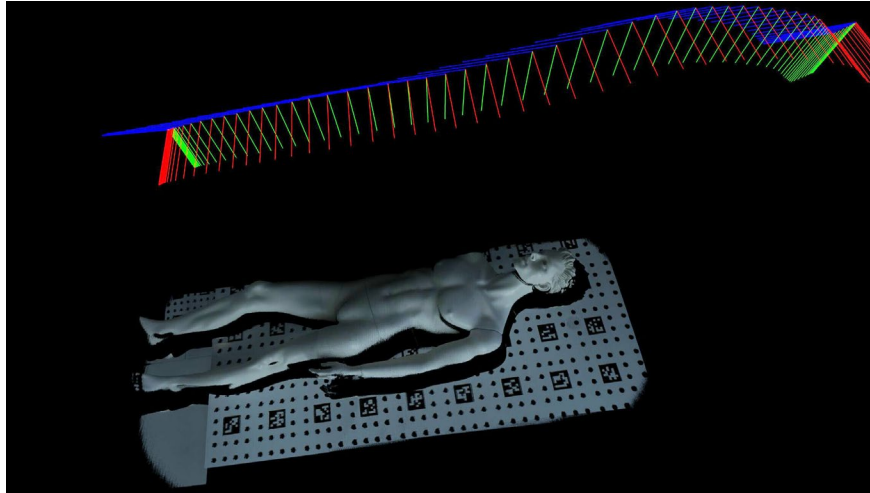
Abb. 5: Ablauf der Registrierung: Der obere Teil der Grafik stellt die Echtzeitberechnung während der Datenaufnahme dar. Der untere Teil beschreibt den Ablauf der Nachverarbeitung.

### 3.5 Erstellung farbiger 3D-Modelle

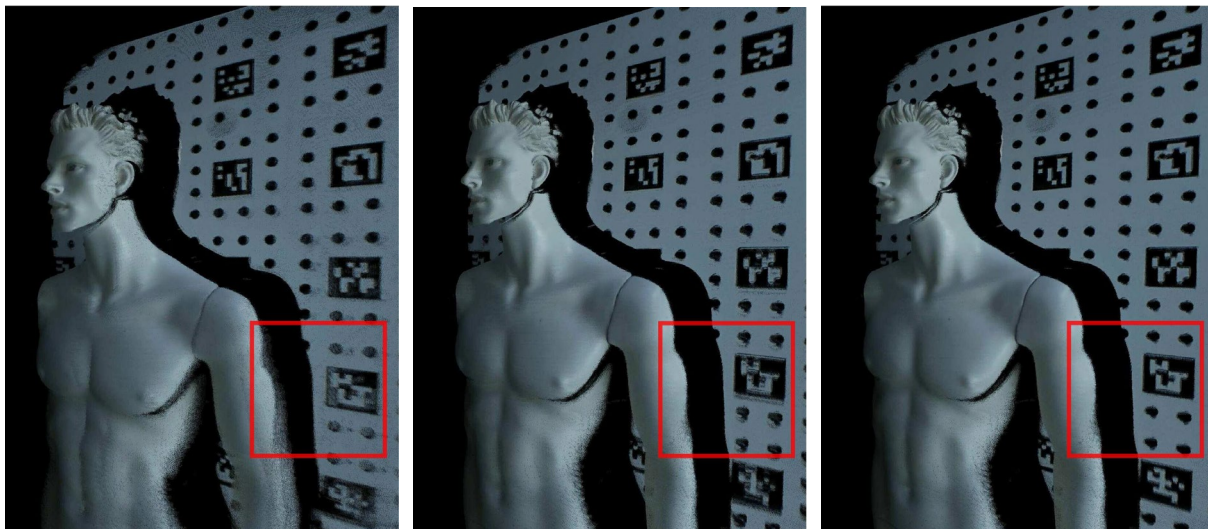
Neben der Bewegungskompensation der 3D-Scans ist die Trajektorien-schätzung eine wesentliche Voraussetzung für die präzise Einfärbung der Punktwolken. Da die Farbbilder zeitversetzt zwischen den Stereobildern der Messkamera aufgenommen werden, besteht zwischen dem 3D-Scan und dem nächsten Farbbild ein Zeitunterschied von bis zu 10 ms. Daher muss die Bewegung des Sensors berücksichtigt werden, um die Farbinformation auf die Punktwolke zu projizieren. Dies wird auf folgende Weise durchgeführt: Zunächst wird die anhand visuelle Odometrie berechnete Trajektorie interpoliert, um die relative Bewegung der Kamera zwischen der Aufnahmezeit des 3D-Scans und dem Farbbild zu berechnen. Hierzu werden die Zeitstempel der Sensorinformationen genutzt. Anschließend wird die Punktwolke anhand der Ko-Kalibrierung von Farbkamera und Stereosensor und der geschätzten Kamerabewegung in das Koordinatensystem der Farbkamera transformiert. Schließlich wird die Punktwolke eingefärbt, indem die Messpunkte anhand des Kameramodells in das Farbbild projiziert werden.

Abbildung 6 zeigt ein Beispiel für eine Punktwolke, die mit dem Sensorsystem in Bewegung im Labor aufgenommen wurde. Eine Schaufensterpuppe wurde in Luft vor einer Kalibriertafel aufgenommen. Das obere Bild zeigt hierbei die farbige Punktwolke und eine Visualisierung der einzelnen Sensorposen. Die unteren Bilder zeigen den Effekt der Trajektorie auf das Gesamtmodell. Das untere linke Bild zeigt das initiale Ergebnis anhand der in Echtzeit berechneten Trajektorien-schätzung der visuellen Odometrie. Durch ein rotes Viereck sind Registrierfehler hervorgehoben. Das mittlere Bild zeigt das verbesserte Ergebnis nach der 3D-Registrierung. Die Registrierung

führt zu einer konsistenten Punktwolke des Gesamtmodells. Die Farbzurordnung ist allerdings ungenau, da hier anhand des zeitlich nächsten Farbbildes eingefärbt wurde, ohne die Bewegung zu berücksichtigen. Dieser Fehler wird im Endergebnis, das im Bild unten rechts zu sehen ist, reduziert. Hier wurde die Zuordnung der Farbinformation anhand der bestimmten Sensortrajektorie verbessert.



Eingefärbte Punktwolke und Sensorposen



Visuelle und inertielle Odometrie

3D-Registrierung

Farbprojektion mit  
Bewegungskompensation

Abb. 6: Punktwolke einer Schaufensterpuppe die durch Bewegung des Sensorsystems im Labor aufgenommen wurde. Oben: Eingefärbte Punktwolke und Visualisierung der Sensorposen. Unten links: Initiales Ergebnis anhand VIO. Unten Mitte: Ergebnis nach 3D-Registrierung. Unten rechts: Verbesserte Farbprojektion anhand interpolierter Trajektorie.

## 4 Ergebnisse

Im Verlauf der Sensorentwicklung wurden verschiedene Tests unter Wasser durchgeführt. Neben der beispielhaften 3D-Erfassung von Testobjekten wurden Genauigkeitsuntersuchungen an Kugelstäben und Ebenheitsnormalen durchgeführt, um den zufälligen und systematischen Messfehler abschätzen zu können.

Zunächst erfolgten diese Tests in Wasserbecken, um den logistischen Aufwand so gering wie möglich zu halten und die Einflüsse wie Wasserqualität, Messentfernung und Sensorgeschwindigkeit selbst exakt steuern zu können. Nach erfolgreicher Absolvierung der Wasserbeckentests konnte das Sensorsystem einem echten Offshore-Einsatz in der Ostsee mit einem ferngesteuerten Unterwasserfahrzeug unterzogen werden. Bei den Tests im Wasserbecken wurde der Sensor für eine Messentfernung von  $2 \text{ m} \pm 0.5 \text{ m}$ , beim Offshore-Einsatz für eine Messentfernung von  $1.3 \text{ m} \pm 0.3 \text{ m}$  eingestellt und entsprechend kalibriert.

Bei ersten Experimenten im mit klarem Süßwasser gefüllten Wasserbecken wurden zunächst systematischer und zufälliger 3D-Messfehler eines Einzelscans bestimmt. Dazu wurden in Anlehnung an die VDI/VDE-Norm 2634 (VDI/VDE 2008) Längenmessfehler und Ebenheitsfehler sowie das Rauschen der 3D-Messpunkte in Abhängigkeit von der Messentfernung untersucht. Der Sensor wurde an einem Kran befestigt und konnte exakt bzgl. der Testmesskörper (Kugelstab, Ebenheitsnormal) positioniert werden. Die ausführlichen Ergebnisse sind in (Bräuer-Burchardt et al. 2022a) dokumentiert. Es ergab sich eine leichte Drift der Kugeldurchmesser und Längenmesswerte proportional zur Messentfernung von 0.08 %. Dieser Fehler konnte mit einer 3D-Korrekturfunktion minimiert werden. Der Ebenheitsfehler nahm bei großer Messentfernung zu (Bräuer-Burchardt et al. 2022a). Hier konnte aufgrund einer zu geringen Messwertanzahl keine Korrekturfunktion bestimmt werden. Aufgrund der Eigenschaften des verwendeten erweiterten Lochkameramodells (Bräuer-Burchardt et al. 2022b) ist jedoch das Auftreten eines entsprechenden entfernungsabhängigen Fehlers bei der relativ großen Tiefenvariation im erwarteten Bereich.

Bei den Messungen im Seewasser konnten aufgrund der Navigation am Remotely Operated Vehicle (ROV) und der im Vergleich zum Wasserbecken erheblich eingeschränkten Sicht keine exakten Entfernungen zu den Messobjekten angefahren werden. Die Messobjekte wurden im Vorbeifahren gescannt. Eine entfernungsabhängige Genauigkeitsuntersuchung wurde anhand der im Nachhinein bestimmten mittleren Messentfernung vorgenommen. Es konnten Messwerte im Bereich von ca. 1.1 m bis 1.6 m Entfernung bestimmt werden. Auch hier wurde eine entfernungsabhängige Drift von 0.4 % bei der Abstandsmessung beobachtet, die jedoch nicht bei der Bestimmung des Kugeldurchmessers zu verzeichnen war. Als mögliche Ursache kommen systematische Fehler bei der Bestimmung des Kugelradius in Betracht. Der zufällige Längenmessfehler bei Wiederholungsmessungen in gleicher Messentfernung verdoppelte sich in etwa im Vergleich zur Wasserbeckenmessung. Allerdings ist dieser Wert nicht signifikant, da zu wenige Messwerte vorhanden waren und keine Garantie für gleichbleibende äußere Bedingungen gegeben werden kann. Einzelne Messwerte sind in Tabelle 1 zusammengefasst. Die kalibrierte Länge der Kugelhantel beträgt dabei 497,612 mm. Bei Sensor 2 tritt außerdem ein leichter Skalierfehler von knapp einem Prozent auf.

Tab. 1: Zufälliger Längenmessfehler bei Messungen im Wasserbecken (Sensor 1) und offshore in der Ostsee (Sensor 2). Die kalibrierte Länge der Kugelhantel ist 497,612 mm.

Sensor 1	Wasserbecken		Sensor 2	Ostsee	
Entfernung	Messung Kugelabstand	n	Entfernung	Messung Kugelabstand	n
1.54 ± 0.00 m	497.602 ± 0.030 mm	6	1.16 ± 0.02 m	499.655 ± 0.105 mm	10
1.94 ± 0.01 m	497.873 ± 0.040 mm	6	1.28 ± 0.01 m	500.067 ± 0.186 mm	10
2.24 ± 0.00 m	498.144 ± 0.037 mm	5	1.57 ± 0.03 m	501.716 ± 0.167 mm	10

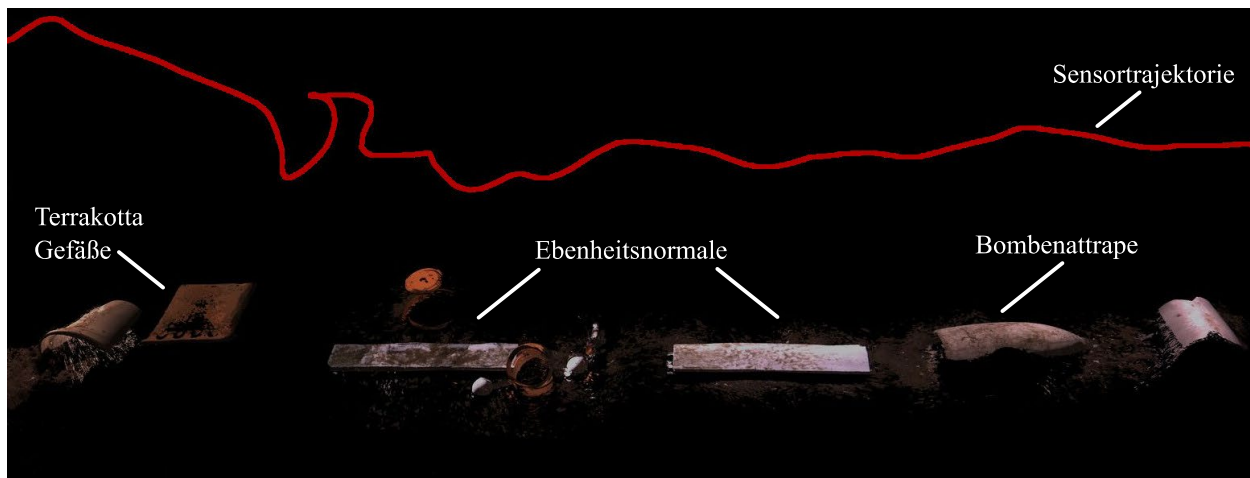


Abb. 7: Beispiele für eine Punktwolke die mit dem Unterwassersensor in der Ostsee in Bewegung aufgenommen wurden. Die rote Linie stellt die bestimmte Sensortrajektorie dar.

Ein signifikanter Ebenheitsfehler wurde bei Entfernungen unter 1.2 m registriert, bei größeren Entfernungen trat kein typischer „Durchbiegungseffekt“ auf. Allerdings kann auch hier wegen zu geringer Anzahl von Messungen in unterschiedlichen Entfernungen keine Korrekturfunktion bestimmt werden.

Insgesamt konnte festgestellt werden, dass sich der zufällige Messfehler bei Seebedingungen im Vergleich zum Wasserbecken in etwa verdoppelte, was sowohl die Wiederholungsmessungen des Kugelabstands als auch das Rauschen der 3D-Messpunkte betrifft. Der systematische Messfehler war bei den Seemessungen wesentlich höher als im Wasserbecken. Ursache hierfür kann eine weniger gute Kalibrierung sein, die jedoch aufgrund fehlender Ressourcen nicht wiederholt und optimiert werden konnte.

Neben der Erfassung der kalibrierten Testkörper wurde das Sensorsystem bei der mobilen Erfassung von Rohrstrukturen validiert. Das Weitern wurden 3D-Scans von einer Bombenattrape und Terrakotta-Gefäßen erstellt. Ein Beispiel für eine in Bewegung aufgezeichnete Punktwolke während der Seerprobung ist in Abbildung 7 zu sehen.

Zur Überprüfung der Trajektorienbestimmung anhand der visuellen und inertialen Odometrie wurden Kugelleisten am Meeresboden ausgebracht. Die Kugeln auf einer Leiste haben einen Durchmesser von 10 cm und einen Abstand von 2,21 m. Die Kugelleisten wurden dann mit dem am

ROV montierten Unterwassersensor eingescannt. In Abbildung 8 ist die zusammengesetzte Punktwolke aus in Bewegung aufgenommenen 3D-Scans abgebildet. Die rote Linie visualisiert hierbei die vom System in Echtzeit geschätzte Trajektorie.

Ein Einzelscan des Sensors erfasst in dieser Aufnahme ein Stück von ca. 40 cm der Kugelleiste. Die Kugelabstände wurden durch ein Einpassen von Kugeln mit festem Durchmesser überprüft. Der mittlere gemessene Kugelabstand ist  $2,212 \pm 0,011$  m. Die mittlere absolute Abweichung der Kugelabstände ist dabei 8 mm. Eine Verbesserung des Ergebnisses durch Registrierung der 3D-Daten konnte in diesem Fall bisher nicht erreicht werden, da die Scans zu wenig 3D-Struktur enthalten. Aufgrund der hohen Reflektivität der Kugeln und Metallprofile konnte Meeresboden und Kugelleisten nicht gleichzeitig in einem Scan erfasst werden.

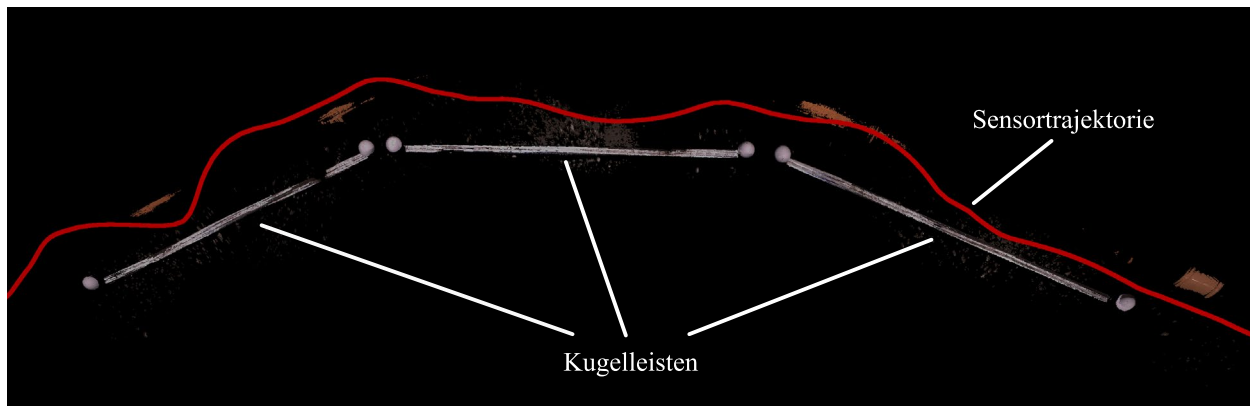


Abb. 8: Zusammengesetzte Punktwolke von drei Kugelleisten aus in Bewegung aufgenommenen Scans mit dem Sensor 2 bei Offshore-Tests in der Ostsee. Die rote Linie visualisiert die Trajektorie des Sensors.

## 5 Fazit & Ausblick

Mit dem im Projekt UWSensor neu entwickelten Sensorsystem konnte eine mobile Erfassung von Strukturen sowohl im Klarwasser als auch bei Experimenten mit einem Unterwasserfahrzeug in der Ostsee demonstriert werden. Die 3D-Rekonstruktion basierend auf aktivem Stereo unterstützt durch Streifenlichtprojektion erlaubt eine flächige Erfassung und eine 3D-Abtastung von bis zu 60Hz. Ein besonderes Augenmerk liegt auf den Kalibrierverfahren für den Unterwasser-Stereo-Scanner, die ausgehend von der geometrischen Modellierung des Strahlverlaufs unter Berücksichtigung der Refraktion an den Medienübergängen und zusätzlichen Verfeinerungsverfahren zu äußerst geringen systematischen Fehlern führen. Das System realisiert eine Bewegungsschätzung sowie eine Fusion der Einzelscans in ein global konsistentes 3D-Modell. Bei mobilen Messungen konnte eine Verbesserung der Punktwolke durch eine automatische Kompensation der Bewegung anhand visueller Odometrie erzielt werden.

In Experimenten im Klarwasser konnten Messergebnisse mit sehr kleinem systematischem Fehler erzielt werden, die im gleichen Genauigkeitsbereich sind wie Messungen in Luft mit vergleichbarem Messvolumen. Bei der Seerprobung zeigte sich eine Verdopplung des zufälligen Längen-



messfehler bei Wiederholungsmessungen in gleicher Messentfernung im Vergleich zur Wasserbckenmessung. Des Weiteren trat ein leichter Skalierfehler von knapp einem Prozent auf. Eine umfassende Auswertung steht noch aus.

Die Anwendung der 3D-Registrierung war nicht auf allen Offshore-Aufnahmen erfolgreich. Aufgrund des eingeschränkten Sichtfelds und Messbereichs ist nicht immer genügend 3D-Struktur vorhanden, um die Registrierung vollständig in allen Raumrichtungen zu beschränken. Hier könnte ein kombiniertes Verfahren, welches Farb- und 3D-Informationen gleichzeitig ausrichtet, die Ergebnisse verbessern. Das vorgestellte Verfahren nutzt zwar sowohl die Farbbilder als auch die 3D-Scans für die Trajektorien-schätzung. Allerdings wird die Trajektorie in einem sequenziellen Verfahren und nicht in einem gemeinsamen Ausgleichsverfahren berechnet. Weitere Forschung ist hier notwendig, um eine hochgenaue und robuste Registrierung in allen Situationen zu gewährleisten.

## 6 Danksagung

Die vorgestellte Arbeit wurde gefördert durch das Bundesministerium für Wirtschaft und Energie im Rahmen des Verbundprojekts UWSensor - Unterwasser-3D-Sensor. Die Autoren bedanken sich bei allen Beteiligten, insbesondere bei Fraunhofer IGD in Rostock, Digital Ocean Lab (DOL) und Baltic Taucherei- und Bergungsbetrieb Rostock, für die Zusammenarbeit bei den Feldtests.

## 7 Literaturverzeichnis

- BLEIER, M., MUNKELT, C., HEINZE, M., BRÄUER-BURCHARDT, C., LAUTERBACH, H. A., VAN DER LUCHT, J. & NÜCHTER, A., 2022: Visuelle Odometrie und SLAM für die Bewegungskompensation und mobile Kartierung mit einem optischen 3D-Unterwassersensor. Photogrammetrie Laserscanning Optische 3D-Messtechnik, Beiträge der Oldenburger 3D-Tage 2022, Jade Hochschule, 394-405.
- BRÄUER-BURCHARDT, C., MUNKELT, C., BLEIER, M., HEINZE, M., GEBHART, I., KÜHMSTEDT, P. & NOTNI, G., 2022a: A new sensor system for accurate 3D surface measurements and modeling of underwater objects. Applied Sciences, **12**(9), 4139, <https://doi.org/10.3390/app12094139>.
- BRÄUER-BURCHARDT, C., MUNKELT, C., HEINZE, M., GEBHART, I., KÜHMSTEDT, P. & NOTNI, G., 2022b: Underwater 3D Measurements with Advanced Camera Modelling. PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science, **90**, 55-67, <https://doi.org/10.1007/s41064-022-00195-y>.
- ELSEBERG, J., BORRMANN, D. & NÜCHTER, A., 2013: Algorithmic solutions for computing accurate maximum likelihood 3D point clouds from mobile laser scanning platforms. Remote Sensing, **5**(11), 5871-5906, <https://doi.org/10.3390/rs5115871>.
- HEIST, S., DIETRICH, P., LANDMANN, M., KÜHMSTEDT, P. & NOTNI, G., 2018: High-speed 3D shape measurement by GOBO projection of aperiodic sinusoidal fringes: A performance analysis. Proceedings of SPIE Dimensional Optical Metrology and Inspection for Practical Applications VII, Vol. 10667, 106670A, SPIE Commercial + Scientific Sensing and Imaging, 17. - 19. April 2018 in Orlando, USA, <http://doi.org/10.1117/12.2304760>.

- HOU, H. & EL-SHEIMY, N., 2003: Inertial sensors errors modeling using Allan varianc. Proceedings of the 16th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS/GNSS 2003), September 2003 in Portland, USA, 2860-2867.
- KRUCK, E., 2014: BINGO 6.6 - Bundle Triangulation for Aerial Photogrammetry and Close Range Applications. <https://bingo-atm.de/solutions/>, letzter Zugriff 31.01.2023.
- SCHÜTZ, M., 2016: Potree: Rendering large point clouds in web browsers. Dissertation, Technische Universität Wien.
- SPRICKERHOF, J., NÜCHTER, A., LINGEMANN, K. & HERTZBERG, J., 2011: A heuristic loop closing technique for large-scale 6D SLAM. *Automatika*, **52**(3), 199-222, <https://doi.org/10.1080/00051144.2011.11828420>.
- QIN, T., LI, P. & SHEN, S., 2018: VINS-Mono: A robust and versatile monocular visual-inertial state estimator. *IEEE Transactions on Robotics*, **34**(4), 1004-1020, <https://doi.org/10.1109/TRO.2018.2853729>.
- VDI/VDE, 2008: VDI/VDE 2634 Optische 3D-Messsysteme. VDI/VDE Richtlinien 2008, Verein Deutscher Ingenieure, Düsseldorf, Germany, 2008, Blatt 1-3.

# Visualisierungs- und Filterungsmethoden von CityGML-Daten in einer VR-Umgebung

HELGE OLBERDING<sup>1</sup>

*Zusammenfassung: Geodaten in VR-Szenen werden vorwiegend beim Einsatz von Game Engines, in vielen Einsatzgebieten möglichst realitätsnah und in einer First-Person-Perspektive, visualisiert. 3D-Stadtmodelle im Format CityGML wurden hierbei eingesetzt, der Import erfolgte mithilfe eines neu entwickelten Plug-ins für die Unreal Engine 5. Für eine gezieltere Informationsvermittlung ist in so einer Visualisierung eine Informationsreduzierung der 3D-Szene notwendig, insbesondere in einer First-Person-Perspektive bei einem interaktiven 3D-Stadtmodell. Die hieraus entwickelte Anwendung demonstriert verschiedene Hervorhebungsmethoden mit Symboliken und Rendering Effekten. Die Hervorhebungstechniken sind wiederum auf weitere Anwendungsszenarien übertragbar.*

## 1 Einleitung

VR-Headsets (Virtual Reality) bzw. Head-Mounted Displays (HMD) gehören zu einer Technologie, welche in den letzten Jahren für die Visualisierung von Geodaten zunehmend an Bedeutung gewonnen hat (EDLER & KERSTEN 2021). Herausforderungen, Schwierigkeiten und Chancen für VR-Headsets werden im Bereich der Geoinformatik in verschiedenen Arbeiten zusammengetragen (ÇÖLTEKIN et al. 2020; VIRTANEN et al. 2020).

Geodaten in VR-Szenen werden in vielen Anwendungsbereichen möglichst realitätsnah visualisiert. Realitätsnahe Visualisierungen werden hierbei vermehrt mithilfe von Game Engines realisiert. Insbesondere für die Orientierung des Nutzens ist eine realitätsnahe Geovisualisierung hilfreich. Die meisten VR-Anwendungen betrachten die 3D-Szene nicht von einer Top-Down-Perspektive, sondern aus der First-Person-Perspektive. Das bedeutet, dass aus der Betrachtungsperspektive mit einer VR-Brille die Gebäude 1:1 in der Größe darstellt, wie sie in der Realität wahrgenommen werden. Dadurch kann ein 3D-Stadtmodell für die Nutzens größer und realitätsnaher wirken. Für eine gezieltere Informationsvermittlung ist in so einer Visualisierung eine Informationsreduzierung bzw. eine Filterung der 3D-Szene notwendig.

Die Informationsreduzierung kann mithilfe von alternativen bzw. abstrakten Visualisierungsansätzen erreicht werden. Das Gegenteil zu einer fotorealistischen Visualisierung ist die nicht fotorealistische Visualisierung (Non-photorealistic Rendering, kurz NPR). Diese Visualisierungen können auch als abstrakte Visualisierungen oder als stilisiertes Rendering bezeichnet werden. Umgesetzt wird eine NPR durch die Anpassung des Renderings, der Texturen und durch den Einsatz von Filtern, welche ihren Ursprung aus der Bildverarbeitung haben. Ein nicht fotorealistisches Rendering nutzt eine Vielzahl von Darstellungen, um entweder die Komplexität in der 3D-Visualisierung zu vermindern, damit ein Kunststil imitiert werden kann oder zur Hervorhebung von

---

<sup>1</sup> Technische Hochschule Würzburg-Schweinfurt, Fakultät Kunststofftechnik und Vermessung, Röntgenring 8, D-97070 Würzburg, E-Mail: Helge.Olberding@thws.de

ausgewählten Objekten. Umgesetzt wird eine NPR durch die Anpassung des Renderings, der Texturen und durch den Einsatz von Filtern, welche ihren Ursprung in der Bildverarbeitung haben (AKENINE-MÖLLER et al. 2018).

Die NPR-Ansätze sollen kontextbezogene Kommunikationen mithilfe abstrakter Methoden verbessern. Die Grundbasis für die meisten entwickelten Konzepte beruht hierbei auf der Kartosemiotik als Stilmittel. Die dargestellten Informationen und Objekte werden hierzu auf das Notwendigste reduziert. Abhängig von der Wirkung der Visualisierung werden Filterung, Texturen oder das Rendering an unterschiedlichen Punkten angepasst (SEMMO et al. 2015).

Hervorhebungsmethoden werden für 3D-Objekte benötigt, um diese potenziell auszuwählen, zu platzieren, zu verschieben oder einfach anzuzeigen. Dabei können die Hervorhebungen in ihrer Darstellung variieren (TRAPP et al. 2011).

Eine Kombination realistischer und alternativer Visualisierungsmethoden kann die Informationsvermittlung fördern und gleichzeitig können Nutzende in der VR-Szene im Maßstab 1:1 navigieren. Hierbei ist es sinnvoll, bestehende Hervorhebungsmethoden im VR-Kontext zu testen und potenzielle neue Methoden zu entwickeln.

Durch die Möglichkeiten der Unreal-Engine, der VR-Brille und dem CityGML-Plug-in ergeben sich verschiedene Forschungsfragen. Welche Möglichkeiten der Interaktionen sind mit CityGML-Daten in der Unreal Engine 5 möglich? Inwieweit bieten sich Bewegungseingaben für das Verändern der Modelle an? Welche Auswirkung hat die Perspektive einer VR-Brille auf das 3D-Stadtmodell?

## 2 Methoden

Für die Visualisierungsvarianten von Stadtmodellen in VR-Anwendungen wurde die Game Engine Unreal Engine eingesetzt. Für die 3D-Szenen wurden exemplarisch freizugängliche Daten der Stadt Berlin und der Stadt Würzburg verwendet, hierzu gehören Metadaten, digitale Geländemodelle und 3D-Stadtmodelle im Format CityGML. Anschließend wurden verschiedene Hervorhebungsmethoden und Interaktionsmöglichkeiten auf Basis unterschiedlicher Techniken entwickelt.

### 2.1 Grundlagen für die VR-Anwendung

Eine Game Engine bzw. Spieleentwicklungsumgebung ist ein Softwareframework, welches die Entwicklung von Anwendungen, die eine Echtzeit-Visualisierung benötigen, ermöglicht. Die Engines vereinen mehrere Komponenten, welche in die virtuelle Welt implementiert werden können, sodass diese wiederum in Echtzeit visualisiert werden kann (TRENHOLME & SMITH 2008). Für interaktive 3D-Geovisualisierungen in Echtzeit bieten sich verschiedenste Game Engines an. Game Engines besitzen vielfältige Möglichkeiten, detaillierte Landschaften und Modelle darzustellen. Als Game Engine wurde die Unreal Engine 5 (UE5) in der Version 5.0 verwendet. Die UE5 ermöglicht es im Vergleich zu deren Vorgänger der UE4 mehr Polygone und Objekte gleichzeitig darzustellen.

Als VR Hardware wurde als HMD die HTC Vive Pro 2 vom Hersteller HTC eingesetzt. Die HTC Vive Pro 2 beinhaltet neben dem HMD noch zwei Motion Controller. Die VR-Brille und deren

Motion Controller werden automatisch von jedem Unreal Engine Projekt erkannt. Das ist ein Vorteil bei der Entwicklung, da die Brille konstant als Kontrollbildschirm für die Anwendung verwendet werden kann.

Für eine neue Anwendung bzw. Entwicklungsumgebung wird ein neues Projekt in der UE5 angelegt. Echtzeit-Anwendungen werden mit einer digitalen Kamera betrachtet. Nutzende lenken die Kamera, um den Bildausschnitt der 3D-Szene zu verändern. Bei VR-Anwendungen wird die Kamera Position und Ausrichtung mithilfe der Kopfbewegungen angepasst.

Eine Visualisierung wird am häufigsten mit Begriffen wie der Realitätsnähe, Echtheit, Glaubwürdigkeit und dem Fotorealismus verbunden. Von diesen Begriffen erfüllt der Begriff der Realitätsnähe am ehesten die Ziele der meisten Geovisualisierungen. Grundlegend sollte eine Visualisierung bestimmten Prinzipien folgen. Sie sollte einen repräsentativen Charakter wie typische und wichtige Merkmale der jeweiligen Objekte besitzen. Daneben sollten die Bedeutung und die Details der Visualisierung eindeutig erkennbar sein. Darüber hinaus sollte die Genauigkeit der Geoinformationen feststellbar sein, damit die Visualisierung eine Berechtigungsgrundlage besitzt eine Geovisualisierung zu sein. Ebenso sollte die Genauigkeit die ursprüngliche Gestalt des Objekts wiedergeben (MACH & PETSCHKE 2006).

Die Unreal Engine 5 hat automatisiert in jedem Projekt einen realistischen Himmel in Form einer Skybox, mit volumetrischen Wolken und einem anpassbaren Sonnenstand. Für die realitätsnahen Visualisierungselemente wurde auf kostenfreie Asset-Bibliotheken zurückgegriffen. Quixel beinhaltet hierzu Physically Based Rendering Material mit hochauflösenden Texturen, insbesondere für die Darstellung der Vegetation.

## 2.2 CityGML-Plug-in

3D-Stadtmodelle liegen in Deutschland vermehrt im Format CityGML vor. CityGML ist ein Anwendungsschema für die Speicherung und den Austausch von Geoinformationen. Der Zweck des CityGML-Standards und des einheitlichen Objektkatalogs ist die Festlegung der Erfassung, Datenhaltung, Bereitstellung und Visualisierung von 3D-Stadtmodellen, damit diese in unterschiedlichen Bereichen und Anwendungsszenarien verwendet werden können (COORS et al. 2016). CityGML ist ein quelloffenes Datenmodell, baut auf einem XML-Format auf und basiert auf den Standards der ISO 19xx Familie (GRÖGER et al. 2012). Ein weiterer wichtiger Kernbestandteil von CityGML ist die Verwendung verschiedener, klar definierter Detaillierungsstufen.

Ein direktes Importieren von CityGML-Daten in die Unreal Engine ist nicht möglich. Es wird eine zusätzliche Software benötigt, um das CityGML-Format umzuwandeln, dabei können je nach Format Informationen wie z.B. die enthaltenen Metadaten verloren gehen. Mithilfe eines neu entwickelten Plug-ins ist es möglich, CityGML-Daten mitsamt ihren Metadaten in die Unreal Engine 5 zu importieren. Das kostenfreie Plug-in ist in Zusammenarbeit der THWS Würzburg, der Abteilung Geovisualisierung, und der Uni Würzburg, der Abteilung Game Engineering, entstanden. CityGML Plug-in ist auf Basis von C++ Code. Der Ordner vom Plug-in wird in das Plug-in Verzeichnis der Unreal Engine 5.0 geschoben.

Nach dem Import wird es in einem Unreal Projekt manuell aktiviert. Anschließend ist es möglich, CityGML-Daten per Drag-and-Drop in das jeweilige Unreal-Projekt zu integrieren. Anschließend kann die importierte Datei mithilfe des Befehls „Build Mesh“ generiert werden. Es entsteht für

jedes Gebäude ein eigenes 3D-Modell bzw. ein Static Mesh mit drei automatisierten einfarbigen Texturen für Boden, Wand und Dachfläche.

Metadaten bleiben bei diesem Verfahren für jedes Gebäude bestehen. Zusätzliche Informationen werden direkt in den jeweiligen 3D-Modellen, unter „Building Data“ hinterlegt. Diese Informationen können manuell angepasst und mithilfe von Blueprints abgefragt werden. Hierbei sind es 21 Felder mit festen Bezeichnungen für die Informationen. Unter die Informationen fallen z.B. das Jahr der Erbauung, die Adresse, die Anzahl der Stockwerke oder das LoD. Weitere Informationen, die dem Modell zugewiesen sind, werden unter dem Abschnitt „Unused Data“ abgelegt, hierzu gehören auch die Maße jeder einzelnen Fläche eines Gebäudes.

Mithilfe des Plug-ins ist es möglich, Gebäude bis zur LoD 2 (Level of Detail) fehlerfrei zu erstellen. Mit dem Plug-in wurde getestet, dass mehrere 10.000 Gebäude in ein Projekt implementiert dargestellt werden können (Abb. 1).

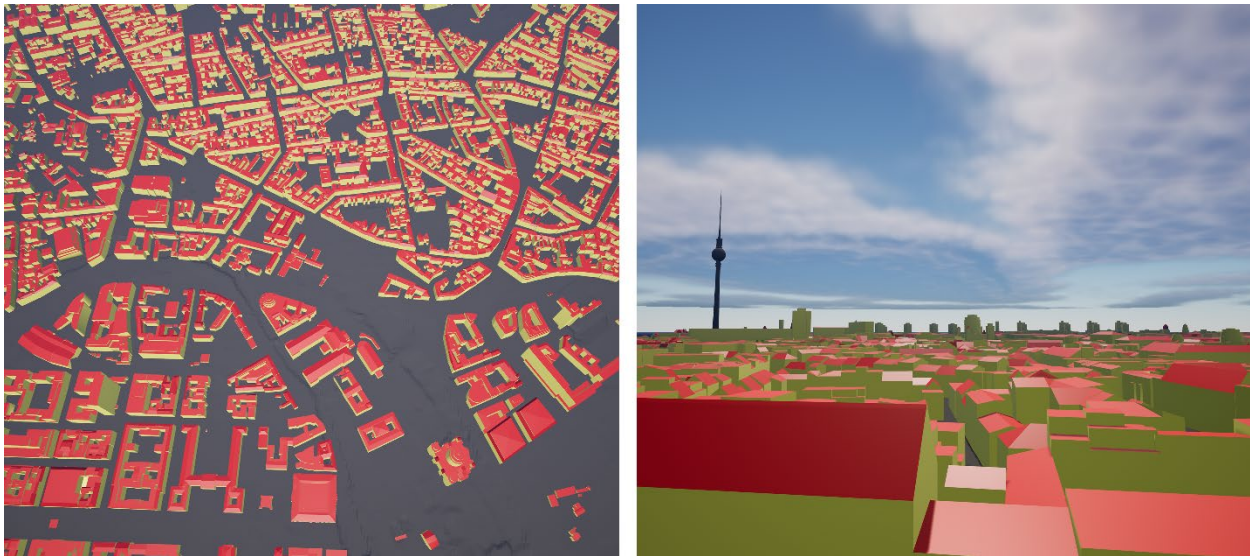


Abb. 1: Generierte Gebäude von der Stadt Berlin, mithilfe des CityGML-Plug-ins

## 2.3 Hervorhebungsmethoden

Es wurde zwischen drei Techniken zur Hervorhebung unterschieden. Zum einen können Objekte hervorgehoben werden durch die Implementierung von Symboliken in Form von Objekten an dem festgelegten Objekt. Diese Symbole können Marker mit 2D-Grafiken oder 3D-Modelle sein. Es sind alternativ auch Highlight-Effekte auf Basis von Partikelsysteme möglich. Mithilfe von Renderingtechniken wie dem Post Processing ist ebenfalls machbar, ausgewählte Bereiche oder die ganze 3D-Szene hervorzuheben bzw. zu verändern.

### 2.3.1 Post Processing

Das Post Processing ist die Nachbearbeitung von digitalen Bildern. Es ermöglicht, die gesamte Darstellung einer Szene im letzten Rendering Schritt zu verändern. In der Unreal Engine können verschiedenste Visualisierungen und Effekte zu dieser Thematik implementiert werden. Die Effekte werden hierbei immer dann angewendet, wenn sich die Kamera innerhalb eines PostProces-

sVolumes befindet. Die Begrenzung des PostProcessVolume kann gelöst werden und die vollständige Szene beeinflussen. Mittels des PostProcessVolume können verschiedenste Parameter angepasst werden. Hierzu zählen zum einen simulierte Effekte physikalischer Kameraeigenschaften wie z.B. Tiefenunschärfe und Linsenreflexionen, aber auch Kantenglättung oder eine Farbkorrektur sind möglich (EPIC GAMES INC. 2022).

Des Weiteren können individuelle Post Processing Materialien genutzt werden, um zusätzliche Effekte zu implementieren. Hierbei können Mapping Verfahren genutzt werden, wie bei der Texturierung einzelner Objekte. Dies findet alles in der Unreal Engine 5 in Form von Materialien statt. Mit den Materialien ist es möglich, NPR Darstellungen zusätzlich zur realitätsnahen Darstellung zu integrieren, dies können z.B. Cartoon-Shader bzw. Toon-Rendering sein. Kartenähnliche Darstellungsmethoden, insbesondere auf Basis von 2D-Karten können auf Cartoon-Shadern basieren. Ein Beispiel ist hierbei eine schwarze Kontur, die als eine zusätzliche Outline für 3D-Modelle dient. Mithilfe von Outlines können 3D-Objekte klarer voneinander unterschieden werden.

Für die Erstellung eines solchen Materials, z.B. für die Erkennung von geometrischen Kanten und Ränder wird die Funktion SceneDepth genutzt. Die Funktion ermöglicht es, die Entfernung einzelner Objekte mit Farbinformationen darzustellen, abhängig von der Szenenposition der Kamera. Durch die Implementierung eines Filters kann die benötigte Distanz festgelegt werden, in der eine Kante dargestellt wird. Durch eine Rundung auf 0 oder 1 werden nur zwei Farbwerte ausgegeben. Alle Werte unterhalb der gewünschten Distanz werden weiß dargestellt, ansonsten schwarz. Im verwendeten Material müssen nun alle weiteren Szeneninformationen integriert sein und anschließend werden diese mit den schwarzen Linien vereint. Das Ergebnis zeigt, durch die Randbetonung, eine deutlich abstraktere zeichnerische Darstellung, in der angrenzende Gebäude deutlich leichter unterscheidbar sind.

Der Vorteil durch eine visuelle Veränderung der Szene durch Post Processing Einstellungen und Materialien ist zum einen die Vielfalt unterschiedlicher Darstellungsmethoden der gleichen Szene, unabhängig von jeglichem vorliegendem 3D-Objekt. Außerdem lassen sich Post Processing Effekte leicht in Interaktionen implementieren.

In der UE5 kann für jedes Mesh der Ausdruck SceneDepth aktiviert werden. SceneDepth gibt die vorhandene Szenentiefe für das jeweilige Mesh aus. Dabei kann SceneDepth die Tiefe an jeder beliebigen Stelle abtasten, auch wenn z.B. andere Objekte sich vor dem Mesh befinden. Mithilfe der Funktion SceneDepth und einer Anpassung der Post Processing Materialien, ist es dadurch möglich, nur auf bestimmte 3D-Modelle Filter einzusetzen. Dadurch ist es möglich, Effekte wie eine Variante eines Röntgenblicks zu visualisieren. Hierzu bietet sich die Outline-Visualisierung an.

### 2.3.2 Kartographische Symbole

Die eingesetzten Symbole basieren auf kartografischen Signaturen. Die Signaturen sind grafische Zeichen, die die Lage, Art sowie qualitativen und quantitativen Merkmale des Objektes durch eine symbolische oder ikonische Darstellung zum Ausdruck bringen. Gleiche Signaturformen (Kreis, Quadrat, Dreieck etc.) werden von Karte zu Karte immer wieder an andere Begriffe gebunden. Hierbei wurde sich zunächst auf die punkthafte Klassifizierung mit zwei technischen Varianten konzentriert.

Die erste Variante besitzt als Grundlage 2D-Objekte. Diese 2D-Objekte beinhalten PNG-Grafiken. Damit eine 2D-Grafik in einem 3D-Raum am besten erkennbar ist, muss diese frontal betrachtet werden. Um zu gewährleisten, dass eine frontale Sicht gegeben ist, wird die Ausrichtung der Symbole dynamisch zu der Kameraperspektive stetig neu angepasst. Dieser Ansatz wurde schon in frühen 3D-Spielen der 90er eingesetzt wie z.B. Doom oder Duke Nukem 3D. Dadurch ist es ein rechenschonender Ansatz, der es ermöglicht aus einer 2D-Karte Signaturen zu nutzen, ohne diese aufwendig für den 3D-Raum neu zu modellieren. Die 2D-Symbole können zusätzlich aus mehreren PNG-Grafiken bestehen, die zusammen eine Sprite-Animation darstellen können.

Die zweite Variante besitzt als Grundlage 3D-Objekte. Hierzu wird ein FBX-Modell mit dem jeweiligen Symbol verknüpft. Neben komplexeren 3D-Modellen können auch einfache Kugeln oder Quadrate genutzt werden. Das Material der 3D-Objekte beschreibt die farbliche Darstellung. Animationen sind mithilfe einer Rigid Hierarchical Animation oder mit einer Morph Target Animation umsetzbar.

Das Vorhandensein und die Positionen der Symbole beruhen entweder auf der Basis einer Tabelle mit Metadaten oder auf festgelegten Informationen aus den CityGML-Gebäuden. Theoretisch ist es möglich, dass Form und Farbe der Symbole dynamisch veränderbar sind.

### 2.3.3 Partikelsysteme

Ein Partikel-Rendering-System bzw. ein Partikelsystem befasst sich mit der Visualisierung von atmosphärischen Effekten und gehört bei den Rendertechniken zu den Visual Effects. Das Besondere an einem Partikelsystem ist die sehr große Anzahl relativ einfacher Grafikobjekte (Partikel) für die Darstellung von Funken, Rauch, Feuer, Wasser etc. als Partikeleffekte. Alle Darstellungen haben hierbei meist keine eindeutig definierbare, nicht glatte und nicht eindeutig abgrenzte Oberfläche. Durch die einzigartigen Eigenschaften sind eigene Renderingsysteme in Game-Engines implementiert. Die Besonderheiten eines Partikelsystems sind zum einen, dass die Grafikobjekte hierzu meist einfache 2D-Objekte sind, welche aus zwei Dreiecken bestehen, diese werden auch als Quad bezeichnet. Zum anderen können Partikel unterschiedlich animiert werden. Hierbei können Parameter von Bild zu Bild variieren, dazu gehören: Position, Ausrichtung, Größe, Texturkoordinate. Animationen sind hierbei manuell oder prozedural möglich. Außerdem werden einzelne Partikel in vielen Fällen kontinuierlich erschaffen und gelöscht. Ein Emitter ermöglicht im System, die Partikel mit einer benutzerdefinierten Rate zu erzeugen. Wenn Partikel sich außerhalb einer festgelegten Zone bewegen oder ihre Zeitspanne abläuft, werden sie wieder entfernt (GREGORY 2019).

Partikelsysteme können z.B. eingesetzt werden um Wasser oder Smogausbreitungen zu visualisieren. Partikelsysteme sind aber auch hilfreich, um einem 3D-Objekt zusätzliche Highlights zu geben, hierbei müssen Partikelsysteme nicht durchgehend aktiv sein. Einzelne Effekte können für einen kurzen Moment abgespielt werden, um den Blick des Nutzens auf bestimmte Punkte zu lenken. Das besondere bei Partikeln ist die Möglichkeit tausende von kleinen animierten Objekten darzustellen.



## 2.4 Interaktive Elemente

In virtuellen Welten werden dem visuellen Sinn Bewegungen vorgetäuscht die, den Gleichgewichtssinn betreffend, nicht exakt der Realität entsprechen. Die Bewegungskontrolle wurde daher innerhalb der Anwendung eingeschränkt, um der Motion Sickness vorzubeugen. Teleportation ist eine alternative Fortbewegungsvariante, durch die keine Bewegungsvisualisierung dargestellt wird. Daraus ergibt sich, dass kein Konflikt zwischen sensorischem und vestibulärem System entsteht. Die Person kann sich mithilfe einer Teleportation durch die 3D-Szene bewegen oder sich an festen Punkten teleportieren. Durch die festen Teleportationspunkte können leichter größere Distanzen überwunden werden und Personen können an einer idealen bzw. gelenkten Position befinden, um bestimmte Informationen besser zu erkennen.

Die Motion Controller können die Bewegungen der Handpositionen abbilden. In den meisten Anwendungen mit VR-Brillen ist das 3D-Modell von Händen oder der Controller das Einzige, was Nutzende in der virtuellen Welt von sich selbst sehen. Mit diesen „Händen“ ist es für viele Menschen verständlich, auf Objekte zu zeigen oder auf Objekte zu drücken, um diese zu aktivieren oder auszulösen. Dementsprechend sind alle interaktiven Elemente auf diese zwei Varianten ausgerichtet.

Informationen können neben der 3D-Visualisierung, mithilfe einer grafischen Benutzeroberfläche (Graphical User Interface, Abk. GUI) dargestellt werden. Hierbei werden 2D-Elemente zu den 3D-Visualisierungen eingebunden. Hierzu gibt es in der Unreal Engine 5 die Head-up Displays (HUD). Ein HUD liefert zusätzliche Informationen z.B. in Form von Anzeigen, Balken, Texten und Bildern. Innerhalb der Anwendung sind im HUD Informationen dargestellt zum Szenario, Koordinateninformationen des Nutzenden, die festen Teleportationspunkte und Angaben, welche Hervorhebungstechniken aktiv sind.

Ein HUD kann je nach Einbindung die Immersion und die Übersicht beeinträchtigen, dies geschieht, wenn sich zu viele Elemente auf der Bildschirmoberfläche befinden. Außerdem ist es schwerer HUD-Elemente in den Randbereichen zu sehen, wenn ein VR-Headset verwendet wird. Im Gegensatz zu einer Desktopanwendung kann sich der Kopf mit einem HMD, nicht anders ausrichten, um Randbereiche der Bildausschnitte auf den Linsen des VR-Headsets zu betrachten. Dementsprechend gibt es zwei Lösungen. Die Erstere ist das Verschieben der HUD-Elemente stärker Richtung Mitte. Die Alternative ist die Erstellung eines HUD-Elements als Objekt in der 3D-Szene.

Durch das implementieren von HUD-Elementen, als Objekte ist es möglich, Interaktionen mit den Motion Controller zu ermöglichen. Damit nicht an einer festen Position das HUD steht, wurde dies für alle Anwendungsszenarien an das 3D-Modell des linken Motion Controllers gekoppelt. Um zu verhindern, dass das HUD die ganze Zeit eingeblendet ist, ist es zunächst unsichtbar. Das HUD wird erst angezeigt, wenn der linke Motion Controller zu einem festgelegten Winkel der Kamera ausgerichtet ist. Für einen Nutzenden ergibt sich nun der Eindruck, als würde eine interaktive Informationsfläche jedes Mal auftauchen, wenn die linke Hand vor dem Kopf gehalten wird. Im HUD haben die Nutzenden nun die Möglichkeit, eine Auswahl an Hervorhebungsmöglichkeiten zu aktivieren (Abb. 2).



Abb. 2: Beispielhafte HUD-Oberfläche in VR

### 3 Ergebnisse

Um die Forschungsfragen zu beantworten, ist eine Anwendung entstanden. Diese Anwendung demonstriert verschiedene Filterungs- bzw. Hervorhebungsmethoden. Dies ist zum Beispiel durch das Ausblenden oder auch das farbliche oder stilisierte Hervorheben ausgewählter Objekte möglich.

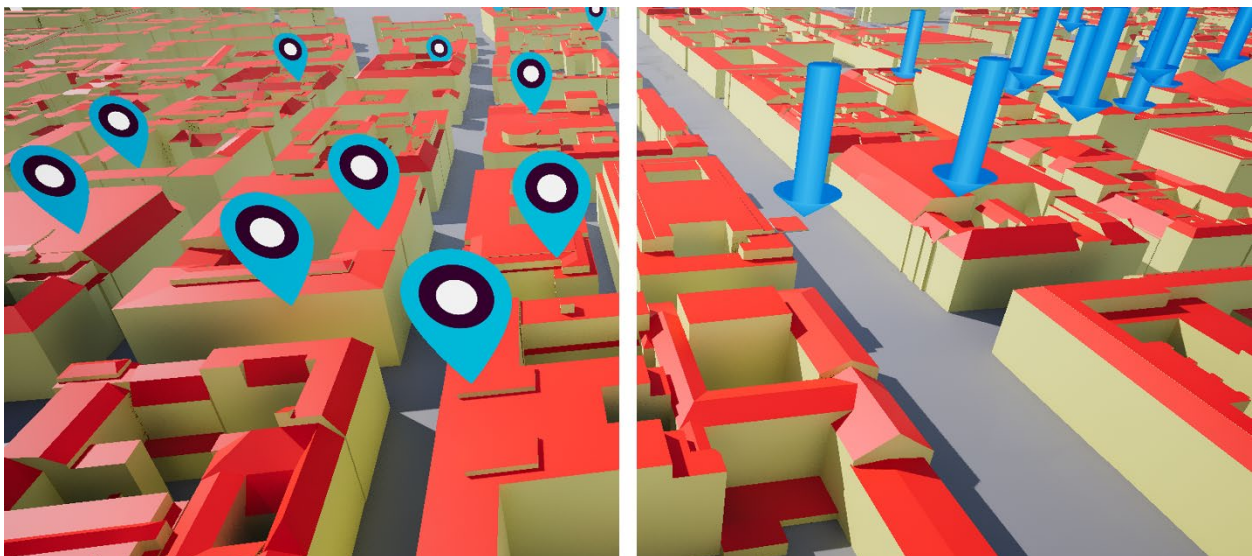


Abb. 3: Verschiedene Symbol zur Hervorhebung selektierter Gebäude

Mithilfe des CityGML-Plug-ins ist es möglich, Gebäude abhängig von enthaltenen Metadaten, in den „Building Data“, hervorzuheben. Hierzu werden Gebäude mithilfe von Blueprints ausgewählt und anschließend werden Symbole (Abb.3) oder NPR-Effekte für diese Gebäude erzeugt. Bei einer

Materialanpassung innerhalb der CityGML-Gebäude können jedoch noch fehlerhafte Visualisierungen auftreten. Mit der Hilfe des HUDs und zusätzlichen Bewegungseingaben können die Hervorhebungen in ihrer Darstellung von Nutzenden, innerhalb der Anwendung, dynamisch ohne Schwierigkeiten verändert werden.

Ein großer Vorteil von allen Hervorhebungsmethoden ist, dass diese Darstellungen einfach miteinander ausgetauscht werden können. Bei den Symbolen und den Partikeln muss nur das Asset und bei dem Post Processing das Material gewechselt werden. Dadurch sind Anpassungen der Hervorhebungen unkompliziert machbar und auf verschiedene Anwendungsszenarien beim Vorliegen der Metadaten und Assets übertragbar. Praxisbezogene Anwendungsszenarien könnten für die Hervorhebung bzw. Filterung zum Beispiel Informationsanzeigen für die Bereiche Tourismus oder dem Energieverbrauch sein. Eine spielerische Hervorhebungsmethode ist ein Röntgenblick, um ein gesuchtes Objekt hinter anderen Gebäuden dennoch erkennbar zu machen. Solche Hervorhebungsmethoden sind insbesondere in der First-Person Perspektive durch die eingeschränkte Übersicht hilfreich (Abb. 4).

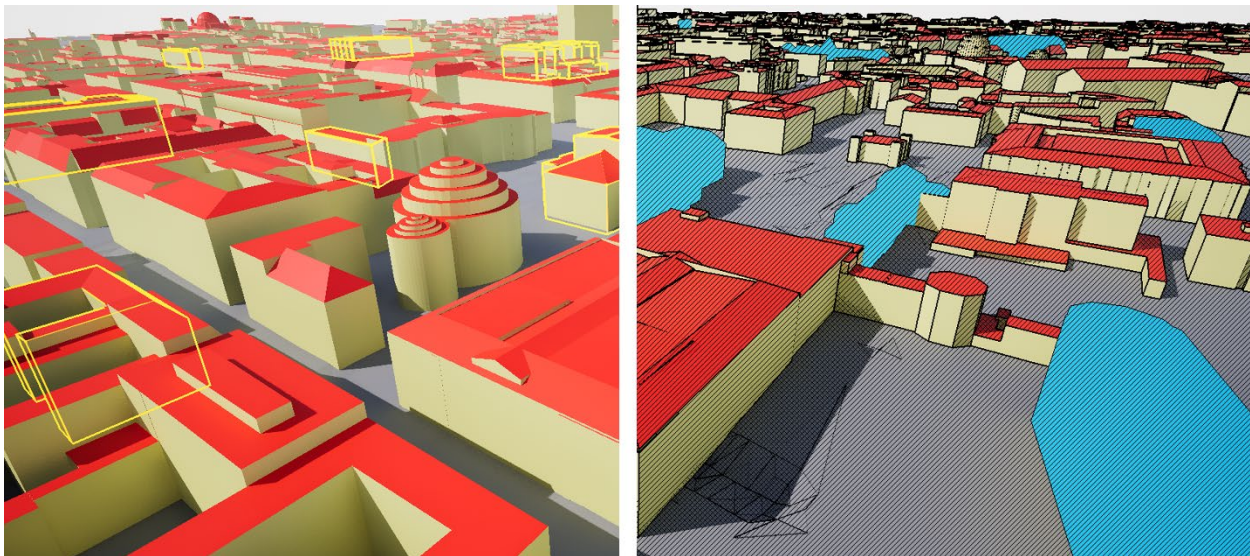


Abb. 4: Vollständige und selektierte Filterungsbeispiele mithilfe von NPR

## 4 Fazit & Ausblick

3D-Stadtmodellldaten werden je nach Anwendungsbereich und Ziel visualisiert. Innerhalb einer Game Engine, existieren kaum technische Limitationen wie die Echtzeit-Renderings aussehen können. Fotorealistische und nicht fotorealistische Visualisierungen können in einer Anwendung sein und mit einem Knopfdruck vom Nutzenden verändert werden. Komplexe dynamische Materialien und Partikeleffekte in Kombination mit einer VR-Brille könnten bei einer genaueren Untersuchung noch mehr sinnvolle Optionen ermöglichen, den Blick des Nutzenden zu lenken.

Durch das schnelle und vollständige Importieren von CityGML-Daten in die Unreal Engine 5 können größere Stadtmodellszenen als ergänzendes Element in Anwendungen implementiert werden. Die Anwendungsszenarien sind erweiterbar und die Hervorhebungstechniken sind mit anderen Metadaten durch das Anpassen von Parametern übertragbar und veränderbar.

Wichtig ist beim Einsatz von der VR-Brille als Ausgabegerät inwieweit die neue Betrachtungsweise spezielle Anpassungen erfordert, um Informationen effektiv zu vermitteln. Die Betrachtung eines Stadtmodells aus der First-Person Perspektive ist im Vergleich zu den meisten Geovisualisierungen auf einem Desktop Bildschirm eine andere. Geovisualisierungen mit dem Ausgabegerät Desktopbildschirm, nehmen ehe eine Top-Down-Perspektive ein. Natürlich könnte ein Stadtmodell auch von oben in VR betrachtet werden, es ist aber auch möglich, vermehrt Ansätze der Hervorhebung aus dem Bereich der Videospiele zu implementieren.

Die beiden größten Features der Unreal Engine 5 sind Lumen und Nanite, welche noch realitätsnähere Visualisierungen ermöglichen. Lumen dient zur Echtzeit Global Illumination und Reflektion. Nanite automatisiert die Skalierung der Polygonanzahl von Objekten abhängig der Kameraposition. Diese waren in der Version 5.0 für VR-Anwendungen noch nicht zugänglich, erst mit der Version 5.1. Daraus ergibt sich eine Anpassung des CityGML-Plug-ins auf die Version 5.1.

Die visuellen Elemente der Anwendung können in nachfolgenden Implementierungen durch das Involvieren zusätzlicher Sinneseindrücke erweitert werden. Dies sind zum einen auditive Elemente, wie z.B. der Einsatz von Sound Cues, um die Betrachtungsrichtung zu lenken, zum anderen haptische Elemente beim Motion Controller, wie z.B. starke und schwache Vibrationen oder unterschiedliche Intervalle.

Eine weitere Aufgabe ist die Untersuchung der Verbesserungsqualität der Informationsübermittlung mithilfe der verschiedenen Hervorhebungstechniken z.B. durch den Einsatz von Eye Tracking. Eye Tracking Möglichkeiten sind hierzu in der HTC Vive Pro 2 integriert.

## 5 Literaturverzeichnis

- AKENINE-MÖLLER, T., HAINES, E. & HOFFMANN, N., 2018: Real-Time Rendering, Fourth Edition. 4<sup>th</sup> ed. Milton: Chapman and Hall/CRC, 651-656.
- CÖLTEKIN, A., LOCHHEAD, I., MADDEN, M., CHRISTOPHE, S., DEVAUX, A., PETTIT, C., LOCK, O., SHUKLA, S., HERMAN, L., STACHOŇ, Z., KUBÍČEK, P., SNOPOKOVÁ, D., BERNARDES, S. & HEDLEY, N., 2020: Extended Reality in Spatial Sciences: A Review of Research Challenges and Future Directions. *IJGI*, **9**(7), 439, <https://doi.org/10.3390/ijgi9070439>.
- COORS, V., ANDRAE, C. & BÖHM, K., 2016: 3D-Stadtmodelle. Konzepte und Anwendungen mit CityGML. Berlin, Offenbach, Wichmann, 13-14.
- EDLER, D. & KERSTEN, T., 2021: Virtual and Augmented Reality in Spatial Visualization. *J. Cartogr. Geogr. inf.*, **71**(4), 221-222, <https://doi.org/10.1007/s42489-021-00094-z>.
- EPIC GAMES INC., 2022: Unreal Engine. Post Process Effects. Online verfügbar unter <https://docs.unrealengine.com/5.1/en-US/post-process-effects-in-unreal-engine/>, letzter Zugriff 15.01.2023.
- GREGORY, J., 2019: Game engine architecture. Third edition. Boca Raton, London, New York: CRC Press Taylor & Francis Group (An A.K. Peters book), 710-712.
- GRÖGER, G., KOLBE, T., NAGEL, C. & HÄFELE, K., 2012: OGC City Geography Markup Language (CityGML) Encoding Standard. Version 2.0.0, OGC Doc. 12-019. Hrsg. v. Open Geospatial Consortium Inc. Online verfügbar unter <http://www.opengeospatial.org/standards/citygml>, letzter Zugriff 01.11.2022.

- MACH, R. & PETSCHKE, P., 2006: Visualisierung digitaler Gelände- und Landschaftsdaten. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg, 1-4.
- SEMMO, A., TRAPP, M., JOBST, M. & DÖLLNER, J., 2015: Cartography-Oriented Design of 3D Geospatial Information Visualization – Overview and Techniques. *The Cartographic Journal* **52**(2), 95-106, <https://doi.org/10.1080/00087041.2015.1119462>.
- TRAPP, M., BEESK, C., PASEWALDT, S. & DÖLLNER, J., 2011: Interactive Rendering Techniques for Highlighting in 3D Geovirtual Environments. *Advances in 3D Geo-Information Sciences*, Thomas H. Kolbe, Gerhard König und Claus Nagel (Hrsg.), Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg (Lecture Notes in Geoinformation and Cartography), 197-210.
- TRENHOLME, D. & SMITH, S., 2008: Computer game engines for developing first-person virtual environments. In: *Virtual Reality*, **12**(3), 181-187. <https://doi.org/10.1007/s10055-008-0092-z>.
- VIRTANEN, J., JULIN, A., HANDOLIN, H., RANTANEN, T., MAKSIMAINEN, M., HYYPPÄ, J. & HYYPPÄ, H., 2020: Interactive Geo-Information in Virtual Reality – Observations and future challenges. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, **44**(4/W1-2020), 159-165, <https://doi.org/10.5194/isprs-archives-XLIV-4-W1-2020-159-2020>.

# Jena4D – Ein virtuelles Stadtgeschichtsbuch zum Mitmachen

FERDINAND MAIWALD<sup>1,2</sup>, CLEMENS BECK<sup>2</sup> & SANDER MÜNSTER<sup>2</sup>

*Zusammenfassung: Jena4D ist der Versuch Bürgerinnen und Bürger an der Digitalisierung ihrer Heimatstadt mitwirken zu lassen. Dafür wird die webbrowserbasierte mobile VR-Anwendung 4DCity bereitgestellt, die 3D-Stadtmodelle, historische Fotografien und weitere Daten im Stadtraum anzeigen kann. Im Rahmen eines Wettbewerbes bekommen die Nutzenden die Möglichkeit eigene historische Stadtfotografien zu digitalisieren sowie historische Bilder nachzufotografieren (Refotografie). Besonders die Digitalisierung privater Bestände ist dabei ein voller Erfolg. Aus den digitalisierten Daten kann mittels angepasster Structure-from-Motion (SfM) Pipeline ein detailliertes historisches 3D-Stadtmodell erstellt werden, dessen Ergebnisse in die Anwendung transferiert werden können. Da die ermittelten Resultate für die Texturierung der Stadtmodelle noch teilweise zu ungenau sind, soll in Zukunft geprüft werden, ob Refotografie exaktere Ergebnisse liefert.*

## 1 Einleitung

Um stadtgeschichtliche Informationen immersiver sicht- und erlebbar zu machen, sind vierdimensionale Darstellungen von Vorteil (vgl. TVERSKY 2005; MUENSTER 2022). Die vierte Dimension entsteht dabei durch die Verschmelzung von dreidimensionaler Information mit einer zeitlichen Komponente. Während bisherige Ansätze zumeist mit unterschiedlichen Modellen der Zeitschnitte operieren, erfolgt dies in unserer Anwendung durch die zeitveränderliche Texturprojektion historischer Fotografien auf dreidimensionale Geometriemodelle (MUENSTER et al. 2021b).

Die genaue Bestimmung der Kameraorientierungsparameter historischer Fotografien erfordert die Verwendung elaborierter Algorithmen (MAIWALD 2022). Bei einer geringen Anzahl an Aufnahmen sind die Ergebnisse noch ungenau und unzureichend verwendbar für Anwendungen in der Virtual Reality (VR) oder Augmented Reality (AR) (MAIWALD & SARDEMANN 2022).

Deshalb beschäftigt sich das Projekt Jena4D damit, wie historische Fotografien unter Zuhilfenahme einer VR-Umgebung exakter positioniert werden können. Dafür sind vor allem eine ausreichende Menge an historischen Fotos notwendig. Deswegen werden Bürgerinnen und Bürger der Stadt Jena in Rahmen eines Wettbewerbs motiviert, historische Motive nachzufotografieren (sog. Refotografie). Refotografie bezeichnet dabei den Prozess der Fotografie einer Szene, von der bereits zu einem früheren Zeitpunkt eine Aufnahme getätigt wurde. Um das zeitgenössische Bild im Folgenden möglichst genau zu überlagern ist das Ziel, die exakte Kameraposition der historischen Fotografie einzunehmen (SCHAFFLAND & VORNBERGER 2022). Die gestellte Aufgabe kann dabei komplett in der browserbasierten VR-Anwendung 4DCity (<https://4dcity.org>) gelöst werden (Abb. 1). Das gesammelte Wissen wird anschließend öffentlich zugänglich gemacht.

---

<sup>1</sup> TU Dresden, Institut für Photogrammetrie und Fernerkundung, Helmholtzstraße 10, D-01069 Dresden, E-Mail: [ferdinand.maiwald@tu-dresden.de](mailto:ferdinand.maiwald@tu-dresden.de)

<sup>2</sup> FSU Jena, Juniorprofessur für Digital Humanities (Bild/Objekt), Schloßgasse 10, D-07743 Jena, E-Mail: [[ferdinand.maiwald](mailto:ferdinand.maiwald), [clemens.beck](mailto:clemens.beck), [sander.muenster](mailto:sander.muenster)][@uni-jena.de](mailto:ferdinand.maiwald, clemens.beck, sander.muenster@uni-jena.de)

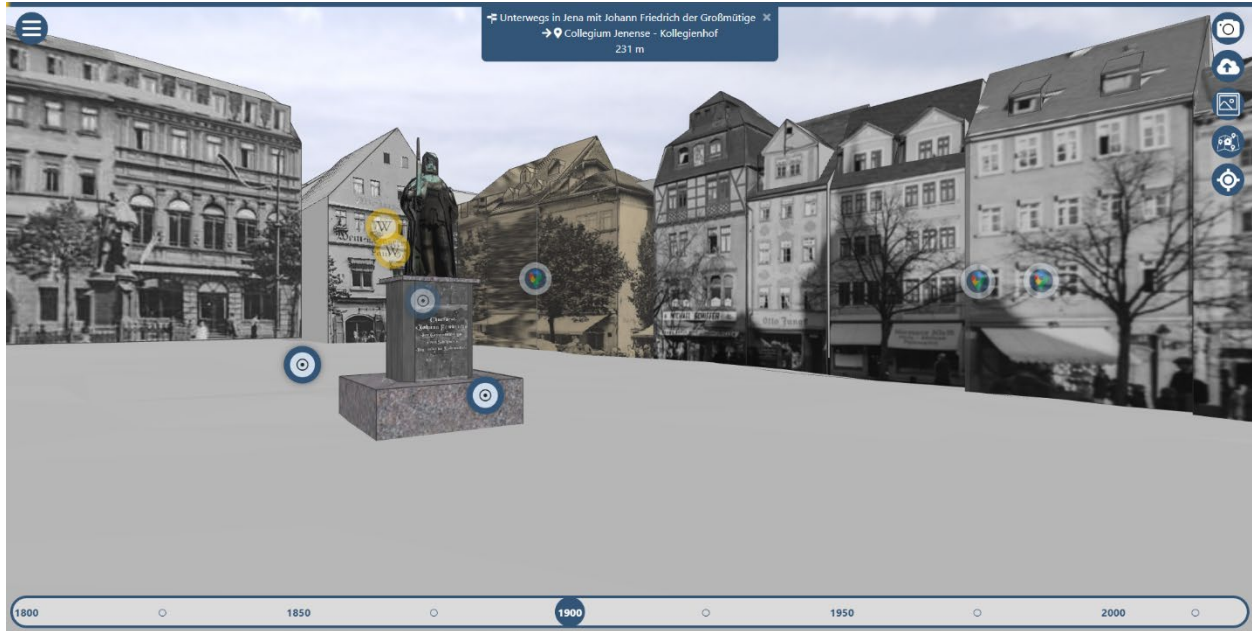


Abb. 1: 4D-Browseranwendung in VR, bei der historische Aufnahmen für die Texturierung dreidimensionaler Gebäudemodelle genutzt werden (abgerufen am 21.11.2022 unter <https://4dcity.org>).

## 2 Technische Basis und Funktionalitäten

Die vorgestellte Anwendung basiert auf zwei Anwendungen, die in vorangegangenen Projekten entwickelt wurden.

Die Struktur der Datenbank stammt aus dem 4D-Browser, der in der Nachwuchsforschergruppe HistStadt4D entwickelt wurde (MAIWALD et al. 2019). Der Kern des Backend ist eine *RESTful API* basierend auf *Node.js* und *Express.js*. Diese verwaltet nicht nur die Datenbankabfragen, sondern ermittelt auch automatisiert die Metadaten aus dem Originalrepositorium.

Die Daten selbst werden entsprechend des CIDOC Conceptual Reference Model (CRM) gespeichert. Daher werden die Daten nicht in einer relationalen Datenbank verwaltet, sondern in einer Graph-Datenbank. Im Falle des 4D-Browsers wird *Neo4j* verwendet (BRUSCHKE & WACKER 2014).

Diese bestehende Datenbank ermöglicht in der darauf aufbauenden VR-Anwendung Jena4D das Hinzufügen von 3D-Modellen, (historischen) Fotografien, digitalen Geländemodellen (DGM), (historischen) Karten und Points of Interest (POI).

Vorentwicklungen zur VR-Anwendung wurden bereits im Projekt Kulturerbe4D geleistet. Um die o.g. 3D-Modelle und Fotografien immersiver darzustellen, wurde eine eigenständige browser-basierte Lösung realisiert. Um die Installation einer App auf den Endgeräten der Nutzenden zu vermeiden, wird auf Grundlage von *Angular* und *three.js* entwickelt. Dies ermöglicht das Aufrufen der Anwendung im Browser sowohl auf dem für die Anwendung empfohlenen Mobilgerät als auch auf dem Desktop-PC. Dabei werden alle gängigen Browser und Betriebssysteme unterstützt.

Durch Nutzerabfrage wird der Zugriff auf die Sensoren des Mobilgerätes realisiert und infolgedessen kann der Nutzerstandort sowie die Orientierung des Gerätes ermittelt werden. Anschließend werden in einem kleinen Radius um den Nutzerstandort automatisiert 3D-Gebäudemodelle

aus OpenStreetMap (meist Level-of-Detail (LOD) 1-2) sowie das nähere digitale Geländemodell über die *Elevation API* geladen. Um städtebauliche Informationen zu erhalten, werden zudem automatisiert von nahegelegenen Gebäuden Informationen aus Wikipedia und weiteren Online-Portalen (Triposo, Yelp etc.) bereitgestellt (Abb. 2).

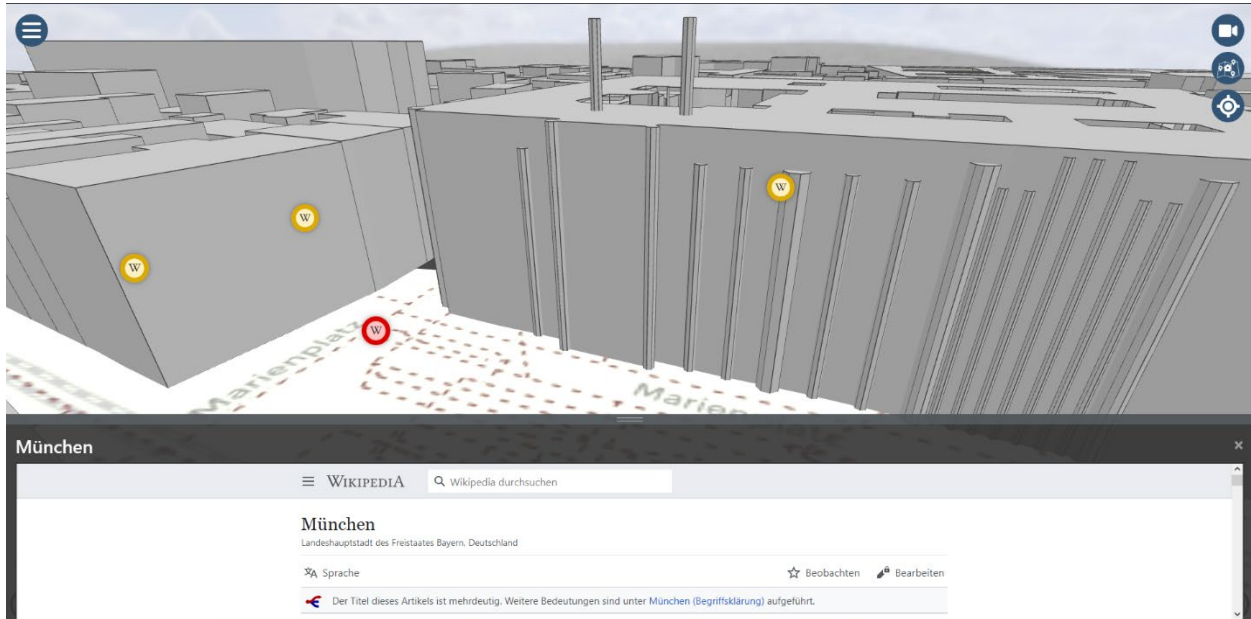


Abb. 2: 3D-Gebäudemodelle aus OpenStreetMap, digitales Geländemodell sowie Gebäude-informationen beim Öffnen der Anwendung in München

Das vorangegangene Projekt Kulturerbe4D konzentriert sich darauf die 3D-Szenen insbesondere in Jena und Dresden mit historischen Fotografien zu füllen. Die Fotografien werden automatisiert mittels eigens entwickeltem Shader direkt auf die Gebäudegeometrie projiziert, um die Gebäudemodelle in Echtzeit zu texturieren. (MÜNSTER et al. 2021a).

Ziel ist es Kulturgeschichte und baugeschichtliche Zusammenhänge in der Stadt zu vermitteln und erlebbar zu machen. Über das Hinzufügen von Stadttouren können Denkmäler und darauf wirkende anthropogene Einflussfaktoren interaktiv erforscht werden.

Im hier vorgestellten Projekt Jena4D wird die entwickelte Applikation als partizipative Wissensplattform zur 4D-Darstellung der Jenaer Innenstadt erweitert.

Dafür werden mehrere neue Funktionalitäten in die Anwendung integriert. Es ist nun möglich eigene historische Fotografien mit entsprechenden Metadaten in die Datenbank hochzuladen. Dafür können analoge Originale mittels der Kamera des Mobilgeräts digitalisiert werden. Um Missbrauch zu vermeiden, wird automatisiert geprüft, ob auf dem digitalen Bild Gebäude bzw. Stadtansichten zu sehen sind.

Des Weiteren wurde ein Refotografie Modul implementiert. Dieses erlaubt die Auswahl einer bereits in der Datenbank gespeicherten historischen Fotografie der Stadt Jena.

Ziel ist es das abgebildete Motiv in der Innenstadt Jenas zu finden. Vor Ort wird dann auf die Kamera des Nutzenden zugegriffen. Mittels transparenter Überlagerung des Originals in der Anwendung gilt es nun das Motiv so präzise wie möglich nachzufotografieren (Abb. 3).



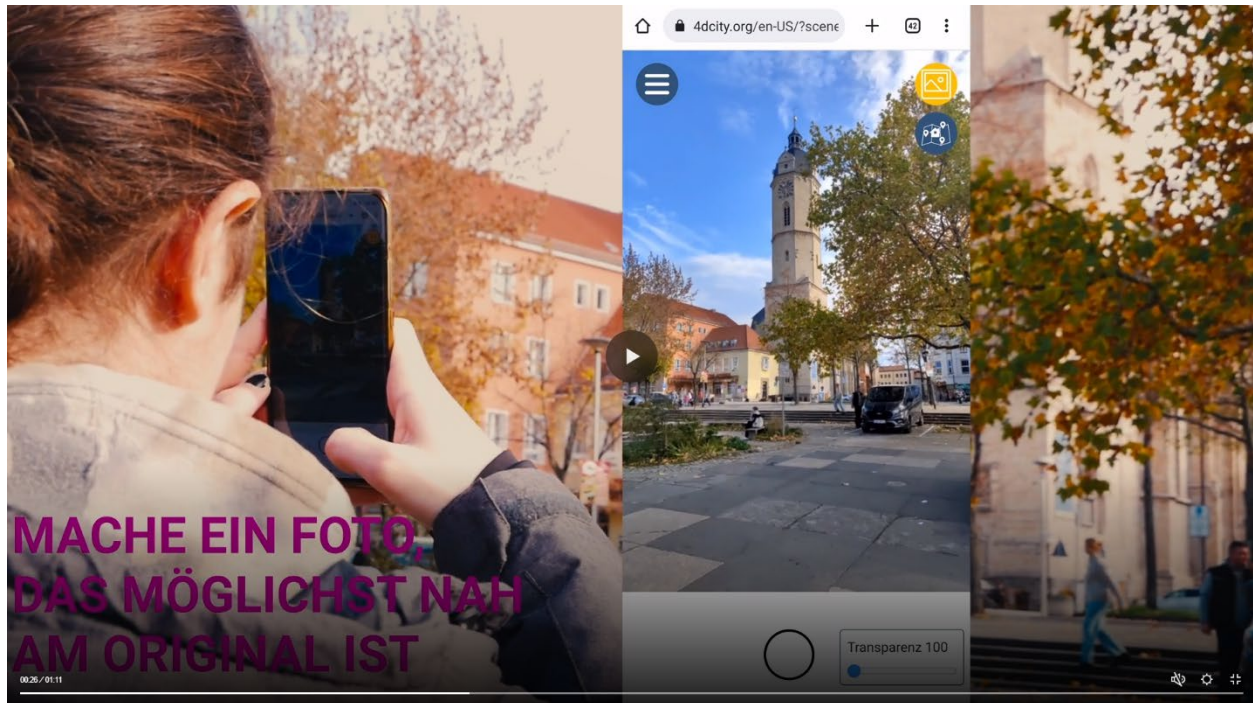


Abb. 3: Ausschnitt eines Erklärvideos zur Nutzung der Refotografie-Komponente der Anwendung im Rahmen eines Bürgerwettbewerbes.

Wenn der Knopf zum Auslösen der Kamera betätigt wird, wird zudem die globale Position sowie die Orientierung des Endgerätes gespeichert und mit übertragen.

Im Folgenden sollen dabei die photogrammetrischen Ansätze, der Bürgerwettbewerb und die Ergebnisse präsentiert werden.

### 3 Automatisierte Berechnung der Kameraparameter historischer Fotografien mittels photogrammetrischer Methoden

Die Bestimmung der Kameraparameter historischer Fotografien werden ausführlich in MAIWALD (2022) behandelt. Die wichtigsten Erkenntnisse und relevanten Weiterentwicklungen werden im Folgenden kurz zusammengefasst.

Historische Stadtfotografien unterscheiden sich stark hinsichtlich ihrer radiometrischen und geometrischen Qualität. Die Bilder sind nicht in homogenen zusammenhängenden Bildverbänden aufgenommen, sondern von verschiedenen Einzelpersonen zu unterschiedlichen Tages- und Jahreszeiten. Dadurch treten starke Unterschiede in Belichtung und Erscheinung der einzelnen Gebäude auf. Zudem können Bildartefakte im Digitalisierungsprozess entstehen.

Dies erschwert die automatisierte Bearbeitung (z.B. in einem Structure-from-Motion (SfM) Arbeitsablauf) enorm. Insbesondere die automatisierte Zuordnung homologer Punkte in historischen Bildern erweist sich als schwierig.

Graph Neural Networks wie SuperGlue (SARLIN et al. 2020) sind seit kurzem in der Lage zuverlässig korrekte Punktkorrespondenzen zwischen historischen Fotografien zu finden (MAIWALD et al. 2021; MORELLI et al. 2022). Dadurch wird ein kompletter SfM-Prozess möglich,

welcher es erlaubt die äußere und innere Orientierung historischer Fotografien automatisiert zu bestimmen. Zudem wird eine historische Punktwolke generiert (Abb. 4).

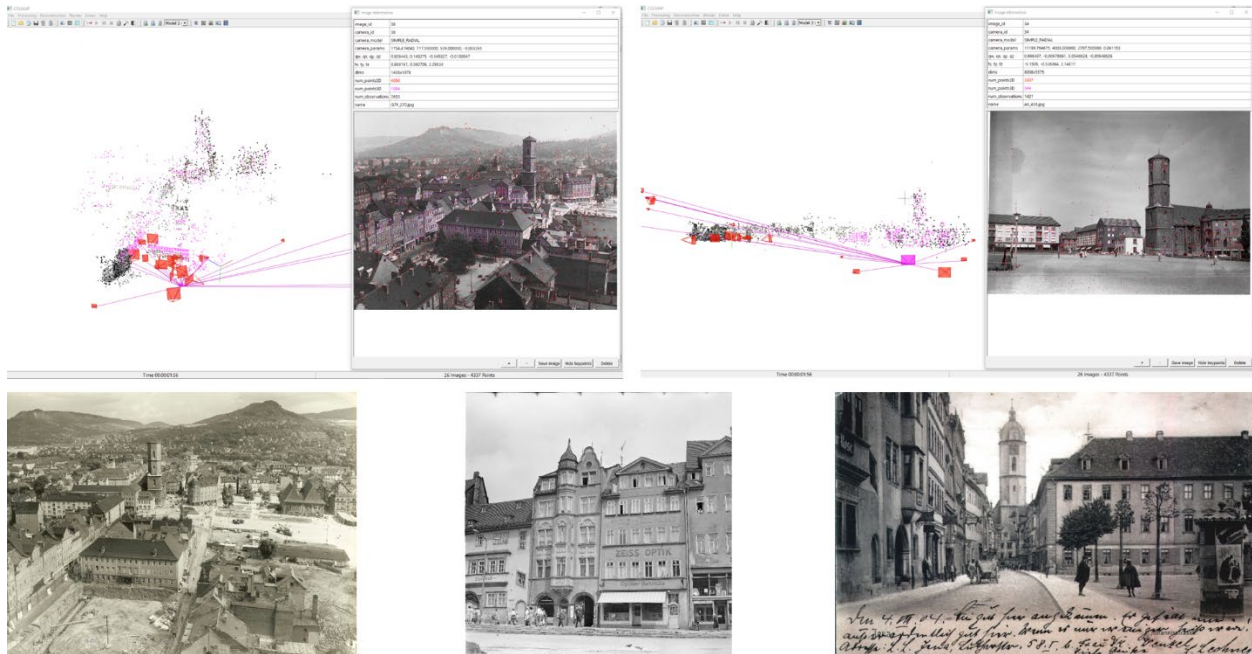


Abb. 4: Oben: SfM-Modell aus historischen Bildern der Stadt Jena welches mittels der Software COLMAP und dem Graph Neural Network SuperGlue erstellt wurde. Unten: Ausgewählte orientierte Fotografien zur Darstellung der radiometrischen und geometrischen Qualität historischer Bilder (MAIWALD 2022).

Ohne Referenzkoordinaten oder Passpunkte, die einen Maßstab festlegen, liegt die historische SfM-Punktwolke und die äußere Orientierung nur im lokalen Koordinatensystem vor.

Durch interaktive Auswahl geeigneter Punkte in der lokalen Punktwolke und im globalen LOD2-Stadtmodell kann eine Helmert-Transformation durchgeführt werden, die es erlaubt die äußere Kameraorientierung in die globalen Koordinaten zu überführen.

Visuell sind die Ergebnisse akzeptabel in der Überlagerung von historischer Fotografie und 3D-Modell, jedoch zeigt die Transformation (je nach Modell) eine Standardabweichung im Dezimeter bzw. kleinen Meter-Bereich, die auch am Modell sichtbar wird (Abb. 5) (MAIWALD 2022). Diese Abweichung ist v.a. durch die geometrische Ungenauigkeit von LOD2-Gebäudemodellen zu erklären. In einem realistischen Szenario liegen der Stadtverwaltung aber oft nur diese Gebäudedaten flächendeckend vor. Laserscanner-Daten sind meist nicht vorhanden.

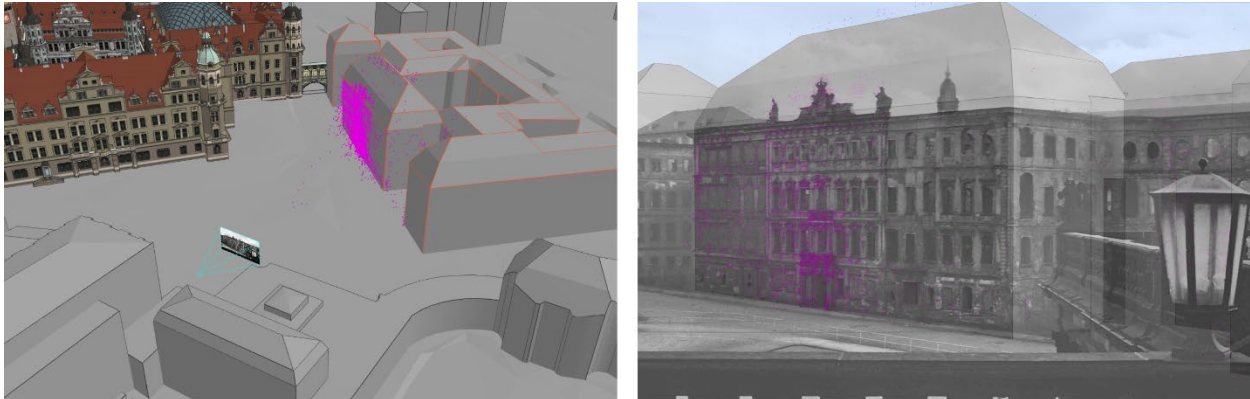


Abb. 5: Links: Visualisierung der Punktwolke im 4D-Browser. Rechts: Überlagerung der Fotografie mit dem 3D-Modell. Es handelt sich dabei um ein Gebäude in Dresden für welches umfassende Experimente durchgeführt wurden (MAIWALD 2022).

Auch die Hinzunahme zeitgenössischer Fotografien mittels kalibrierter Kamera sowie die Referenzmessung von Passpunkten am Gebäude verbessert die geometrische Qualität der äußeren Kameraorientierung nur begrenzt (MAIWALD & SARDEMANN 2022).

Während die Ergebnisse in der 4D-Browser Anwendung zur Überlagerung von Fotografien verwendbar sind, sind diese für eine immersive Darstellung in VR/AR (zur Texturprojektion) noch zu ungenau. Im Folgenden soll u.a. kritisch geprüft werden, ob die Ergebnisse des Bürgerwettbewerbes für eine genauere Überlagerung verwendet werden können.

#### 4 Bürgerwettbewerb “Das schönste Jena aller Zeiten“

Vom 25. Oktober 2022 – 04. Dezember 2022 konnten Jenaer Bürgerinnen und Bürger mit der VR-Anwendung eigene historische Bilder in die Datenbank hochladen sowie die Refotografie-Komponente der Anwendung nutzen, um bei der Lokalisierung historischer Bilder zu unterstützen. Um die Nutzerinnen und Nutzer an das Thema heranzuführen, wird die Anwendung der Refotografie mittels eines Videos erklärt.

Dazu werden 20 Bilder, die zufällig aus der Datenbank ausgewählt werden in der Browser-Anwendung angezeigt. Anschließend kann eines der Fotos auf denen man das Motiv erkennt, ausgewählt werden, um es in der heutigen Zeit nachzufotografieren. Ziel ist es möglichst genau den Aufnahmewinkel zu treffen und anschließend den Auslöser in der Anwendung zu betätigen. Nach Bestätigung wird das Foto mit den zugehörigen Metadaten, gespeicherter Position und Orientierung des Gerätes in der Datenbank abgelegt.

Am Ende des Wettbewerbes wurden über 4000 Bilder digitalisiert sowie 67 Bilder refotografiert. Aufgrund der geringen Anzahl an nachfotografierten Bildern sind die Ergebnisse für die Refotografie-Komponente leider nur teilweise aussagekräftig.

## 5 Ergebnisse

Für ein Konvolut von 3112 Bildern von Jena, die im ersten Wettbewerbszeitraum von Bürgerinnen und Bürger digitalisiert wurden, wird getestet, ob mit der Methode aus MAIWALD (2022) ein SfM-Modell der Innenstadt errechnet werden kann. Dazu werden die Bilder vorgefiltert, um v.a. Innenstadtsichten und Schrägluftbilder zu erhalten (Abb. 6).

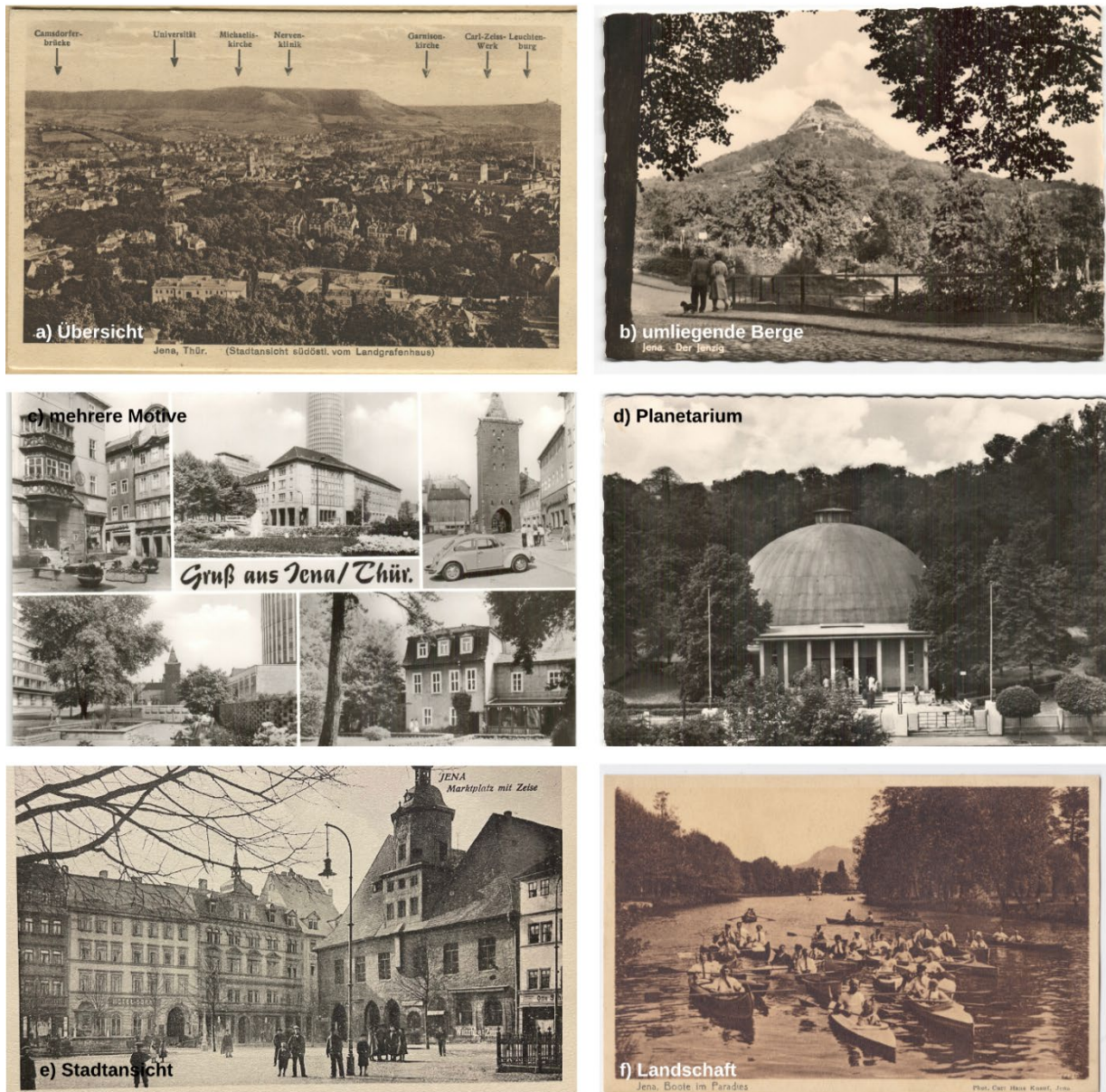


Abb. 6: Beispiel für Bilder, die im Rahmen des Wettbewerbes digitalisiert wurden

Bei diesem Datensatz geschieht das teilautomatisiert. Bilder, die mehrere Motive enthalten (v.a. auf Postkarten) sowie die Rückseiten von Postkarten werden mittels inhaltsbasierter Suche gefiltert (MAIWALD et al. 2021). Bei den ca. 1000 verbleibenden Digitalisaten werden Ansichten der

umliegenden Berge, weitere Landschaftsfotografien, Innenräume (keine urbanen Fotografien) und des Planetariums (als extra SfM-Modell) händisch entfernt.

Es verbleiben 635 Bilder was eine enorme Steigerung zum bereits berechneten Modell mit 26 Bildern (Abb. 4) bietet. Ergebnis der SfM-Prozessierung ist die innere und äußere Kameraorientierung der historischen Fotografien sowie die berechnete Punktwolke (Abb. 7).

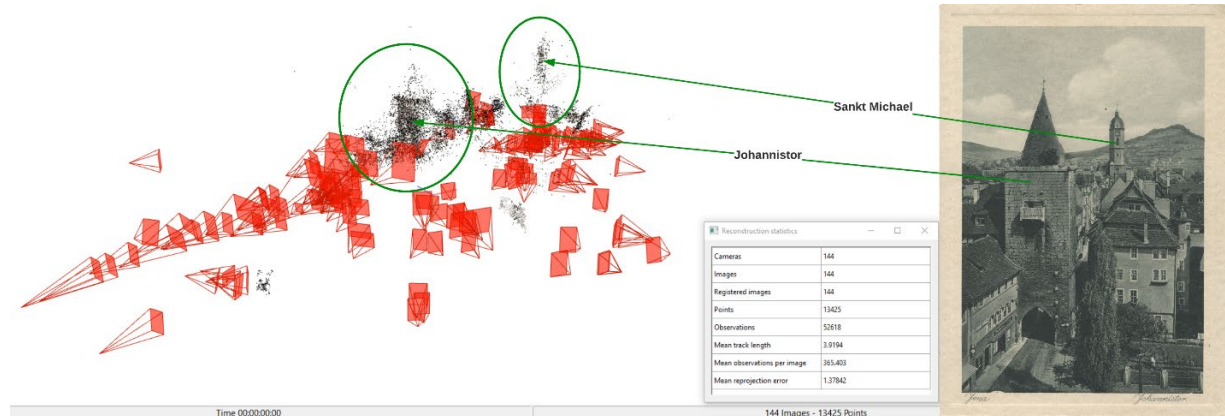


Abb. 7: Größtes berechnetes SfM-Modell der Stadt Jena bestehend aus 144 historischen Fotografien und Postkarten

Das größte berechnete Modell umfasst 144 verortete Fotografien, die v.a. den Innenstadtbereich um das Johannistor abdecken. Zudem berechnet COLMAP 20 weitere Modelle von einzelnen Gebäuden der Innenstadt bestehend aus weniger Fotografien (Abb. 8).

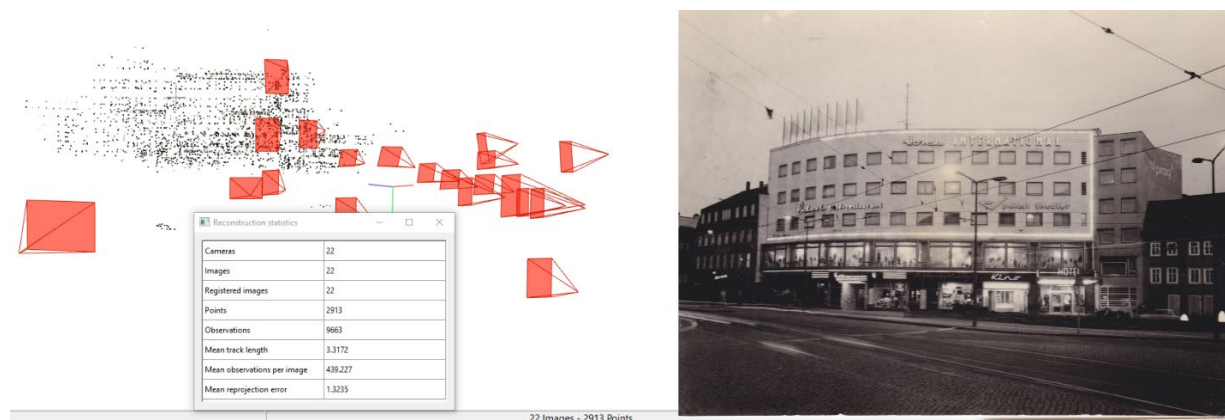


Abb. 8: Weiteres beispielhaftes Modell aller verorteten Bilder des Hotel International in Jena, das zum jetzigen Zeitpunkt noch nicht mit dem Innenstadtmodell verknüpft werden kann

Gegenwärtig können noch nicht alle Modelle miteinander verknüpft werden, da die Ausrichtung der Fotografien im Rahmen der Digitalisierung noch nicht berücksichtigt wurde. Das erschwert die Merkmalsuche, da das gewählte Verfahren nicht rotationsinvariant ist (SARLIN et al. 2020). Dies soll in nachfolgenden Experimenten verbessert werden um ein gesamtes Modell der Stadt zu erhalten.

Des Weiteren soll versucht werden die Bilder, die im Refotografie-Modul aufgenommen wurden mit der 3D-Szene zu überlagern, um einen Vergleich zur bisher erreichten visuellen Genauigkeit

(siehe Abb. 5) zu erhalten. Erste Ergebnisse zeigen die Überlagerung der Refotografie mit dem historischen Original (Abb. 9).

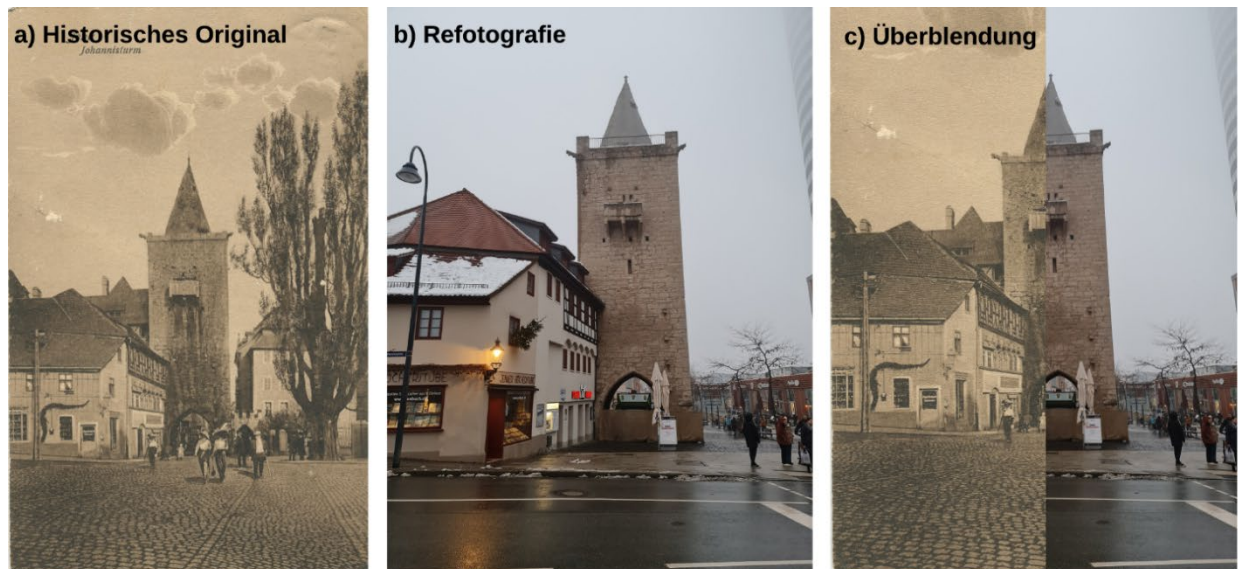


Abb. 9: Refotografie des Johannistors (Mitte) eines historischen Originals (links), welches ein Nutzer im Rahmen des Wettbewerbes erstellt hat. Rechts wird die überblendete Fotografie dargestellt

Die Ermittlung der Kameraparameter und die daraus resultierende Überlagerung in der 3D-Szene sind aufgrund der Verlängerung des Bürgerwettbewerbes noch nicht abgeschlossen. Generell sind auch zu wenige Bilder refotografiert worden, so dass sich noch keine statistischen Aussagen treffen lassen.

## 6 Fazit und Ausblick

Das Projekt Jena4D ist der Versuch Bürgerinnen und Bürger an der Digitalisierung ihrer Heimatstadt mitwirken zu lassen. Über die bereits weit entwickelte und stabil laufende Webanwendung konnten viele neue Nutzerinnen und Nutzer gewonnen werden. Das generelle Feedback zur Anwendung inklusive Stadttouren und Informationen zur Stadtgeschichte ist durchweg positiv. Während die Digitalisierung eigener (privater) analoger Fotografien aufgrund der hochgeladenen Anzahl ein voller Erfolg ist, sind womöglich die technischen Hürden für die Bedienung der Refotografie-Komponente trotz Erklärvideos noch zu hoch. Es hat sich gezeigt, dass die intuitive Benutzung von VR-Anwendungen noch nicht für die breite Masse zugänglich ist. Deshalb besteht immer noch ein Bedarf an lösungsorientierten Strategien, um VR simpler und intuitiver zu gestalten. Nichtsdestotrotz konnten die Digitalisate für photogrammetrische Prozesse verwendet werden. Obwohl diese nicht professionell photogrammetrisch digitalisiert wurden, konnten die Kameraparameter der digitalisierten Fotografien in einem vollautomatischen SfM-Prozess berechnet werden. Aufgrund der geringen Anzahl an Refotografien und die Verlängerung des Wettbewerbes sind die Ergebnisse zur visuellen Genauigkeit leider noch nicht aussagekräftig. In ersten Experimenten zeigen einige Aufnahmen bessere Ergebnisse als im vollautomatischen Prozess während andere deutlich schlechter überlagert sind. Statistische Aussagen darüber sind noch nicht zu treffen. Da das

Refotografie-Modul weiterhin in der Anwendung bestehen bleibt, werden weitere Tests zeigen, ob z.B. eine Mittelung der Kameraorientierungsparameter sinnvoll ist und die Genauigkeit über Nutzerdaten verbessert werden kann.

## 7 Literaturverzeichnis

- BRUSCHKE, J. & WACKER, M., 2014: Application of a Graph Database and Graphical User Interface for the CIDOC CRM. Access and Understanding–Networking in the Digital Era. Session J1. The 2014 annual conference of CIDOC, the International Committee for Documentation of ICOM, 1-3.
- MAIWALD, F., 2022: A window to the past through modern urban environments – Developing a photogrammetric workflow for the orientation parameter estimation of historical images. Technische Universität Dresden, Qucosa.
- MAIWALD, F., BRUSCHKE, J., LEHMANN, C. & NIEBLING, F., 2019: A 4D information system for the exploration of multitemporal images and maps using photogrammetry, web technologies and VR/AR. *Virtual Archaeology Review*, **10**(21), 1-13.
- MAIWALD, F., LEHMANN, C. & LAZARIV, T., 2021: Fully Automated Pose Estimation of Historical Images in the Context of 4D Geographic Information Systems Utilizing Machine Learning Methods. *ISPRS International Journal of Geo-Information*, **10**(11), 748.
- MAIWALD, F. & SARDEMANN, H., 2022: Accuracy Analysis Of Estimated Camera Orientation Parameters Of Historical Images – Case Study of Georg-Schumann-Bau In Dresden (Germany). *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, **43**(B2-2022), 823-829.
- MORELLI, L., BELLAVIA, F., MENNA, F. & REMONDINO, F., 2022: Photogrammetry Now And Then – From Hand-Crafted To Deep-Learning Tie Points. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, **48**(2/W1-2022), 163-170.
- MUENSTER, S., 2022: Digital 3D Technologies for Humanities Research and Education: An Overview. *Applied Sciences*, **12**(5), 2426.
- MUENSTER, S., BRUSCHKE, J., MAIWALD, F. & KLEINER, C., 2021a: Software and Content Design of a Browser-based Mobile 4D VR Application to Explore Historical City Architecture. *Proceedings of the 3rd Workshop on Structuring and Understanding of Multimedia heritAge Contents*, 13-22.
- MUENSTER, S., LEHMANN, C., LAZARIV, T., MAIWALD, F. & KARSTEN, S., 2021b: Toward an Automated Pipeline for a Browser-Based, City-Scale Mobile 4D VR Application Based on Historical Images. *Communications in Computer and Information Science*, 106-128.
- SARLIN, P.-E., DETONE, D., MALISIEWICZ, T. & RABINOVICH, A., 2020: SuperGlue: Learning Feature Matching with Graph Neural Networks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4938-4947.
- SCHAFFLAND, A. & HEIDEMANN, G., 2022: Heritage and Repeat Photography: Techniques, Management, Applications, and Publications. *Heritage*, **5**(4), 4267-4305.
- TVERSKY, B., 2005: Visuospatial Reasoning. *The Cambridge Handbook of Thinking and Reasoning*, 209-240.

# Experiences from the Creation of a Virtual Tour through a Stalactite Cave with VR Gear

PETER BAUER<sup>1</sup> & WERNER LIENHART<sup>1</sup>

*Abstract: This paper describes the 3D survey of an Austrian stalactite cave and the derivation of a Virtual Reality (VR) tour through the data set with the Game Engine Unity. The survey of such a cave beholds many challenges like narrow passages, abstinence of light and scan obstructions in the large stalactite fields.*

*Documenting this huge cave-network required several scanning sessions since the year 2016. The limitations of a cave survey have been overcome by the usage of a broad spectrum of sensors, geodetic total stations, different types of static terrestrial and kinematic laser scanners, SLR cameras and even DJI Mini drones.*

*However, the different surveying techniques deliver varying texture qualities and point accuracies which resulted in inhomogeneous 3D data. Combining these data sets in a common coordinate system and merging them into a single 3D mesh required a solid geodetic reference frame and postprocessing in commercial software products from the laser scanning industry, structure from motion solutions and applications from the entertainment sector.*

*The final 3D model is presented in a VR application as the complexity of the geometry exceeds the capability of the conventional 3D viewers. The user can move along the mesh, has to duck in narrow passages and handles a torch for the exploration of the dark underground. This way students and school pupils get an immersive impression of the dimensions of the cave while they learn about the 3D survey.*

## 1 Introduction

The Katerloch is a major Austrian cave with very diversified terrain properties, as presented in figure 1. It is not only famous for an immense density of stalactites in several stalactite fields throughout its entire structure, but also for its large network of huge domes and narrow passages. The cave's entrance opens up in a huge hall, which leads to a 45m deep shaft that ends in a large underground dome with one of the largest stalactites in Austria. A big part of the cave is accessible for visitors, who can climb down the cave with ladders through the narrow corridors to experience the stalactite fields and even an underground lake that lies 140 m below the entrance hall.

The main challenges for surveying sessions in caves, are the accessibility with equipment (dimensions and weight of the gear), the abstinence of light (camera-based systems) and the limited freedom in the setup design due to the dimensions of the cave and safety reasons. Also, the measurement of the massive number of topological features (stalactites) requires high accuracy and high resolution from the used instruments. Although simple ruler measurements and hand-drawn maps are still best practice in conventional cave exploring, there are already good examples of the usage of laser scanner in underground scenarios like the scanning of the *Eisriesenwelt* in

---

<sup>1</sup> Graz University of Technology, Institute of Engineering Geodesy and Measurement Systems, Steyrergasse 30/II, 8010 Graz, Austria, E-Mail: [peter.bauer, werner.lienhart]@tugraz.at



Austria (BUCHROITHNER & GAISECKER 2020), *Skocjanske* cave in Slovenia (WALTERS & HAJNA 2020) or a Messinian gypsum cave in Italy (DE WAELE et al. 2018).

In comparison to these surveys, there are some features of the survey in the Katerloch that makes this project special. First of all, the solid geodetic reference frame with millimetre accuracy on which the whole survey is based on. The coordinate system was established in 2011 by a geodetic traverse, that was executed with total station measurements throughout the whole structure. Therefore, not only the relative performance of the applied methods could be investigated, moreover the overall accuracy and systematic errors of the surveying techniques could be assessed.

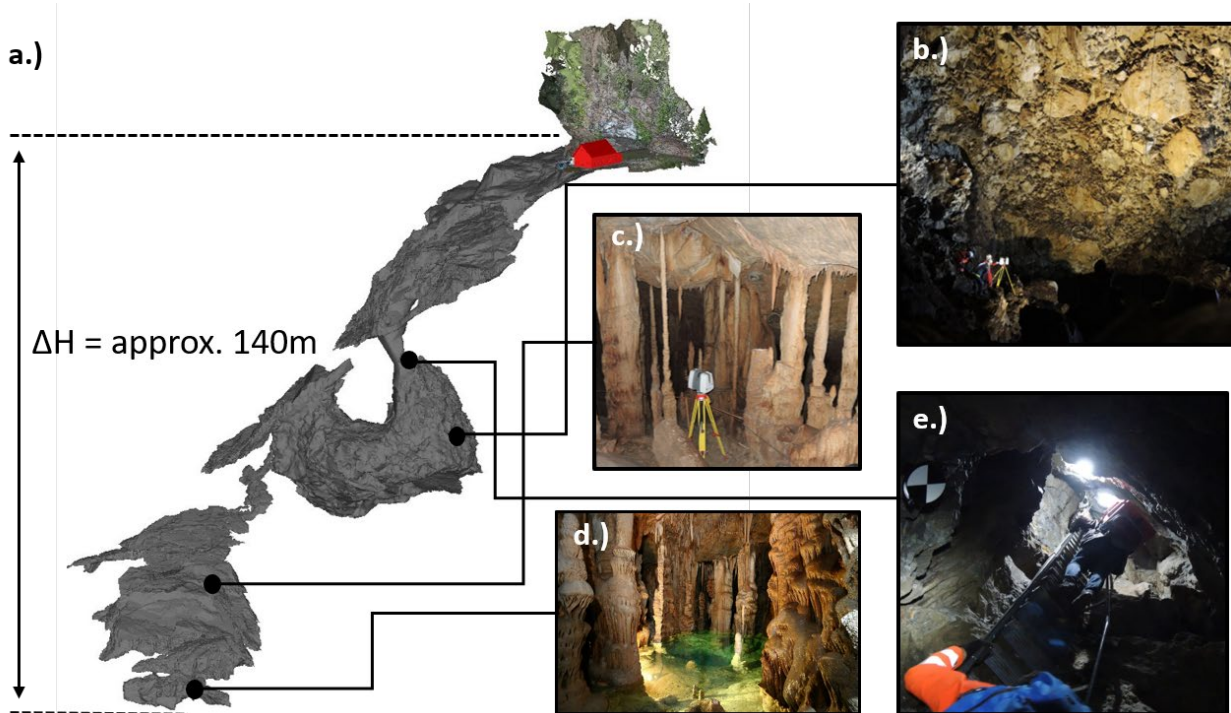


Fig. 1: a.) 3D model of the cave b.) Marteldome c.) stalactite field d.) underground lake e.) corridor with ladders

Furthermore, a big selection of different geodetic sensors has been used. Due to the huge extend of the cave and the challenges of several sections, the effort was too large for a single surveying session. Therefore, since 2016, it has been a training ground for surveying students of Graz University of Technology (TUG) and six measurement campaigns have been executed on a yearly basis. During these campaigns the equipment has been updated and adjusted to the current state of the art in laser scanning. Therefore, the data sets do not only compare different methods but also represent the advances in the technologies over the last years. With the establishment of commercial laser scanners in the late 90s (GEOFFREY 2020) it still can be counted as a relatively new technology, in comparison to tachymetry or photogrammetry. Due to vast changes in the computer industry regarding CPUs and graphic-cards the possibilities of software products for point cloud processing are rapidly increasing. Even in the short period of time since the start of the campaigns in 2016, paradigms in laser scanning have been shifted. With the upcoming of on-board registrations and the increased usage of cloud to cloud approaches in software solutions, the trend is heading towards smaller and mobile solutions like the RTC360 or the BLK series (LEICA

GEOSYSTEMS 2020). The availability of massive high-resolution point clouds has affected also the software processing behaviours of 3D data and geodetic software solutions. Here, a lot has been learned from advances in the entertainment sector regarding texturing of meshes and automatic meshing algorithms.

In order to cover a wide range of settings for the students the main objectives have been changed in every surveying campaign. Special focuses have been placed on the performance of handheld laser scanners, the absolute performance of the registration of static laser scanners in the stalactite fields or the scanning through water to capture the ground level of the underground lake. The investigation of the current state of texture quality provided by all the systems, has been a constant aim through the campaigns and a lot has been learned of what does work in this environment and which approaches have failed.

## 2 From 2D plans to 3D models and gameful design with VR

At the beginning of the 19<sup>th</sup> century (GEISLER 2023), adventurers and explorers have started to document the Katerloch with hand-drawn sketches (figure 2a) and the first 2D maps. The initial aim of the 3D survey was to update these historic maps, to accurately determine the volume of the underground domes and the capture the dimensions of the most prominent stalactites, as it can be seen in figure 2b.

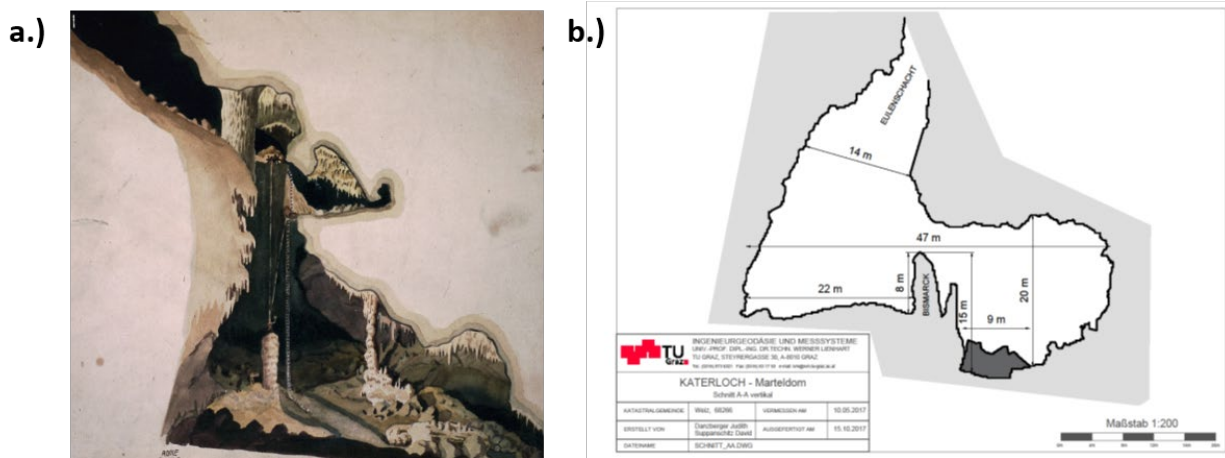


Fig: 2: a.) Sketch from Adolf Mayer 1899 (GEISLER 2023) and b.) plan of the Marteldome from TUG 2017

However, soon it became obvious that 2D plans are no suitable representation of the complex geometry of the cave anymore and that new approaches had to be found for the presentation of the 3D data to the scientific community as well as to students and school pupils.

Beside attempts of 3D printing selected cave parts and producing 3D PDFs as giveaways, the presentation of the data set with Virtual Reality gear was a very promising approach. The Virtual Reality technology originates from the entertainment sector and has been around since the early 1960s. However, the industry received a real boost only in the last decade, when advances in home computing enabled manufacturers like HTC or Oculus to produce VR headsets for the mass market (VIRUTALREALITYSOCIETY 2017). With the establishment of VR devices on the consumer market,

it became also an established tool in the scientific world for 3D data visualisation. The benefit of the technology is the ego-perspective of the user. The user can access the data set from inside and has no complexity limits in visualisation, which exceeds the capability of conventional 3D viewers. The process of integrating video game elements in a non-game context is summarized as Gamification by the definition given by DETERDING et al. (2011). This term has gained popularity in the last years and is often mentioned in the context of digitalisation and technical VR applications. However, for the beneficial adaption of software approaches for technical applications the more precise term gameful design (DETERDING 2015) is more adequate to describe the developed geodetic software presented in this paper.

### 3 Applied surveying methods

When the survey started in 2016 with the Leica P40 scanner [range noise = 0.4 mm @ 10 m, (LEICA GEOSYSTEMS 2016)] it was best practice to register each scan individually with at least four target plates with known coordinates. Therefore, scan setups have been treated like total station setups, which have been registered via free stationing in the Leica Cyclone software. This required a very dense and highly accurate reference frame throughout the whole cave. The reference frame was created in 2011 by a local surveying company and densified by TUG within the following campaigns. The registered P40 data sets from 2016 and 2017 cover the major caverns of the cave and provide a highly accurate and homogenous data set.

Although the free stationing of scanners has been mostly replaced by cloud to cloud approaches and redundant network configurations, these P40 scans are still the most reliable data foundation in the Katerloch data set.

However, the major back-draws of high accuracy scanners like the P40 scanner or other comparable devices in cave surveys, are their weight, their dimensions and their scanning time. This makes them unhandy and hard to operate in narrow passages. For the smaller corridors the BLK360 [3D point uncertainty = 4 mm @ 10 m (LEICA GEOSYSTEMS 2023a)] provided a more mobile and lightweight solution, although the measurement precision is decreased and the noise in the point cloud is increased. Small scanners are mainly designed for agile cloud to cloud matching, which improves the onsite operation, but the uncertainty in the measurement setups accumulates very fast and is prone to systematic deviations without the usage of external reference information. In order to sustain a high accuracy in the overall model the BLK360 setups have been used in combination with the P40 measurements, just to fill gaps in the existing point cloud. The BLK360 setups for each corridor has been co-registered with cloud to cloud-matching, merged together and furthermore transformed as rigid objects onto the existing P40 point clouds. The transformation parameters have been investigated and for most of the parts deviations below one centimetre have been established. This has been also verified at additional terrestrial target signs in the scene. This workflow is demonstrated in figure 3.

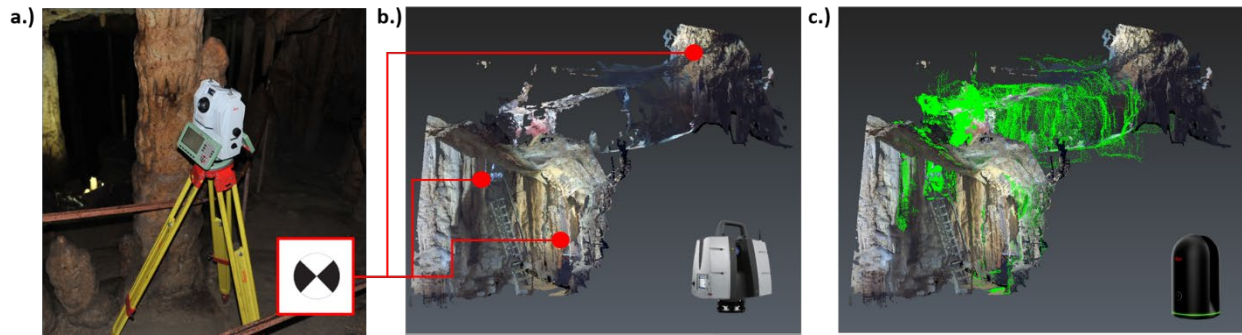


Fig. 3: a.) Leica MS60 for reference measurements, b.) registered P40 point cloud and c.) transformed BLK360 point cloud

Two vertical corridors linking the major stalactite fields are only accessible with ladders, which made also the use of the BLK360 in a tripod setup nearly impossible. These passages have been scanned with the handheld BLK2GO device [range noise = 3 mm; accuracy = 10 mm (LEICA GEOSYSTEMS 2023b)]. The registration of the BLK2GO does not support conventional target signs, because the scan resolution is not sufficient for the detection of these in any software. Therefore, the registration has been also carried out with cloud2cloud-matching of the BLK2GO data onto the existing P40 point cloud. The challenge here was the drift of the BLK2GO in the trajectory of the device, because the implemented GrandSLAM algorithm relies on camera and laser based odometry. For the observation area hardly any light sources were available, therefore the positioning algorithm could only rely on the laser-based slam algorithm and faced a loss in positioning accuracy. It was not possible to manually postprocess the trajectory of the BLK2GO and to optimize the resulting point cloud. To minimize the significant drift in the data set, the observation area has been divided into separate shorter runs. The transformation on the P40 data revealed deviations up to several centimetres for the both corridors (25m and 50m long), which was sufficient for the determination of the passages in the overall model.

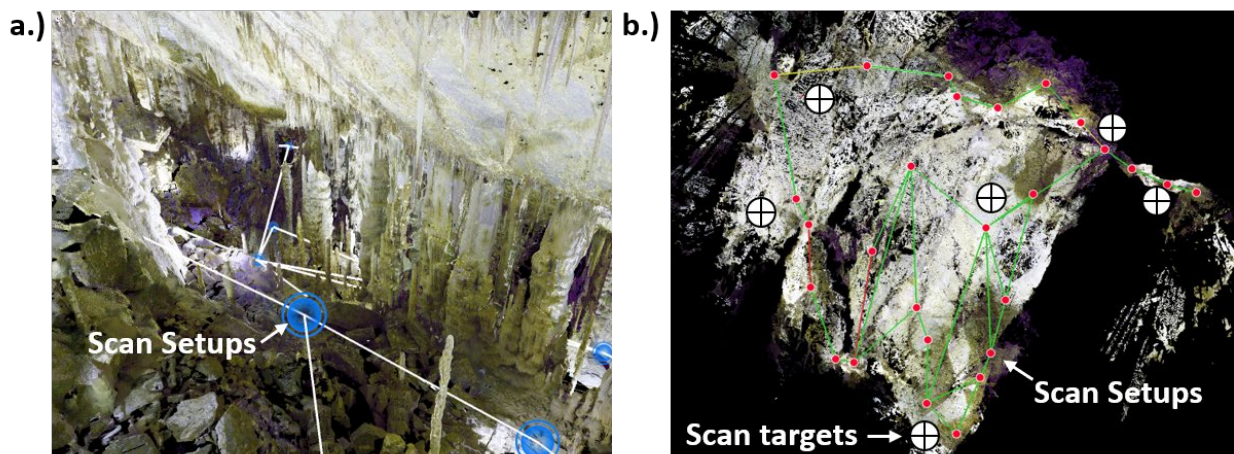


Fig. 4: a.) 3D point cloud with scan setups and b.) 2D network configuration in a stalactite field

The medium range scanner Leica RTC360 [3D point uncertainty = 1.9 mm @ 10 m (LEICA GEOSYSTEMS 2018)] provided a good balance between accuracy and mobility for the capturing of

the huge stalactite fields. The massive number of features required an increased number of setup points and due to obstructions by stalactites an obstacle free line of sight to four target plates was not possible. With the Leica Cyclone Register 360 the point clouds of the RTC360 have been registered on the bases of cc-matching in a redundant network configuration, which is optimised and geo-referenced by the coordinates of the few visible ground control points in the scene. The georeferenced point cloud has been compared to the P40 point cloud in overlapping areas and the differences of a few millimetres were not significant for the survey.

For some selected parts of the cave also structure-from-motion approaches have been tested with small DJI Mini 2 drones and SLR cameras and few ground control markers. The results have been compared to the scanner measurements. Due to the limitations of the pathways, no solid setup geometry could be established in the camera locations and also there was limited possibility for application of ground control markers for supporting the bundle adjustment to preserve the stalactites. Therefore, vast systematic deviations have been revealed in the comparison to the P40 measurement and only laser-based techniques have contributed to the geometrical documentation of the cave.

## 4 3D modelling and mesh creation

The registered point cloud for the Katerloch survey is only an intermediate result. The point cloud can serve for visualisation purposes and software products like the Leica Cloudworx plug-in can easily derive 2D plans out of it. Regarding VR applications a very common approach is a simple fly-through through the raw point cloud. This interaction method is not only of high risk of motion sickness (CHANG & KIM 2020), but also the experience depends strongly on the resolution of the point cloud because the user sees more or less only a fixed wallpaper.

More sophisticated 3D model interactions require the transformation from the raw point cloud into a triangulated mesh. Due to additional topological information of a mesh (e.g. surface normals) it supports 3D printing, complex mathematical operations like volume calculations, more stable deformation analysis with C2M or M2M (HOLST et al. 2016) and better integration into graphic software products, because point clouds are no native data type in the entertainment sector. By meshing the data set the user can move along the mesh in a VR application and interact with the cave in a more immersive manor. The user has to duck to pass narrow corridors or and has to stretch in order to look over an obstacle.

Automatic meshing operations are a basic feature of every point cloud processing software. They have been improved a lot over the last years and such functionally can also be found in open source products like CloudCompare or Meshlab. For the creation of the Katerloch mesh the commercial software 3DReshaper (and the new version: Leica Cyclone 3DR) have been used as they support every processing step, from point cloud filtering, 3D meshing up to mesh cleaning and postprocessing in a single working environment. Different modelling strategies can be found in GEIST et. al. (2018).

For the automatic meshing algorithm, the point cloud has been filtered and resampled to avoid artefacts and degenerated triangles. Blunders in the point cloud caused by visible surveying gear and inattentive crew members have been removed manually in advance. The original resolution of

the point clouds varies from a few millimetres to a few centimetres depending on the sensor type. Therefore, the whole data set has been resampled and smoothed to an average triangle size of 1-2 cm. This was a compromise between performance (size of the mesh) and level of detail. In this configuration the overall mesh resulted in a total size of 5 GB and the shape of the single stalactites was still properly represented, as it can be seen in figure 5.

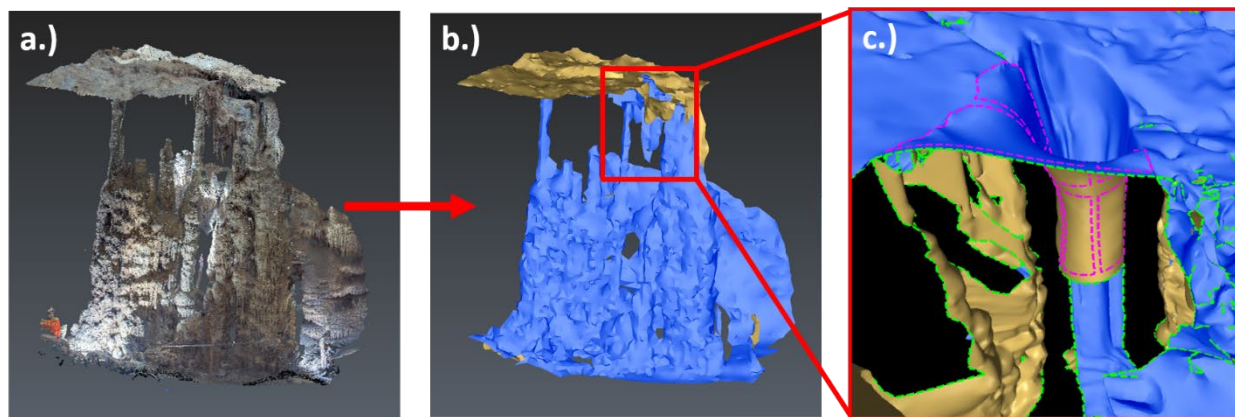


Fig. 5: a.) raw point cloud b.) automatic generated mesh c.) manual hole filling and cleaning process

Defining the triangle size significantly affects the quality of the automatic meshing algorithm, because it indirectly defines the size of holes that are automatically filled and which features are merged. Due to the massive amount of features the decision, if a hole in the mesh is a scan obstruction or separate stalactites, cannot be easily generalized for the whole data set and false classifications can never be avoided, which have to be corrected manually. Therefore, every automatic meshing approach in every state-of-the-art software product is only semi-automatic because the scan data requires a lot of manual cleaning, hole filling and reconstruction of obstructed areas. For the stalactite fields the ratio of fieldwork to postprocessing can be assumed with 1:15, meaning one day of cave scanning will result in three working weeks of mesh modelling. This takes also into account that the few geometric features with sharp edges like stairs, handrails or doors have to be separately modelled within Revit or comparable software products.

## 5 Texture and surface properties

Although the dimensions of the Katerloch would be impressive enough, it is the colourful appearance of the stalactites and the game of light and shadow in the caverns that attracts most visitors. White lime, brown clay and rusty iron ore shine together, when hit by a light source. This is such an important feature and therefore it was always an important aim of the survey to find an appropriate measurement device to document also the different surface properties of the stalactites. However, the complex light situation in the scene, consisting only of few light spots from the tourist paths and scattered torch light, was challenging for the texture quality of the scanning devices. Figure 6 shows the same location in the cave captured with three different devices under comparable circumstances. It can be seen in figure 6, the internal camera of the BLK2GO has no internal brightness correction and the resulting panorama images give the point cloud only a very dark and inappropriate texture. The internal camera of the P40 has a comparable quality, therefore

an external camera system (Istar) has been used for texturing the point cloud at the setup locations. The Istar camera delivers 360° HDR images and it can be seen that the brightness is improved a lot and single features can be easily identified and the colours in the panoramas have a good saturation.

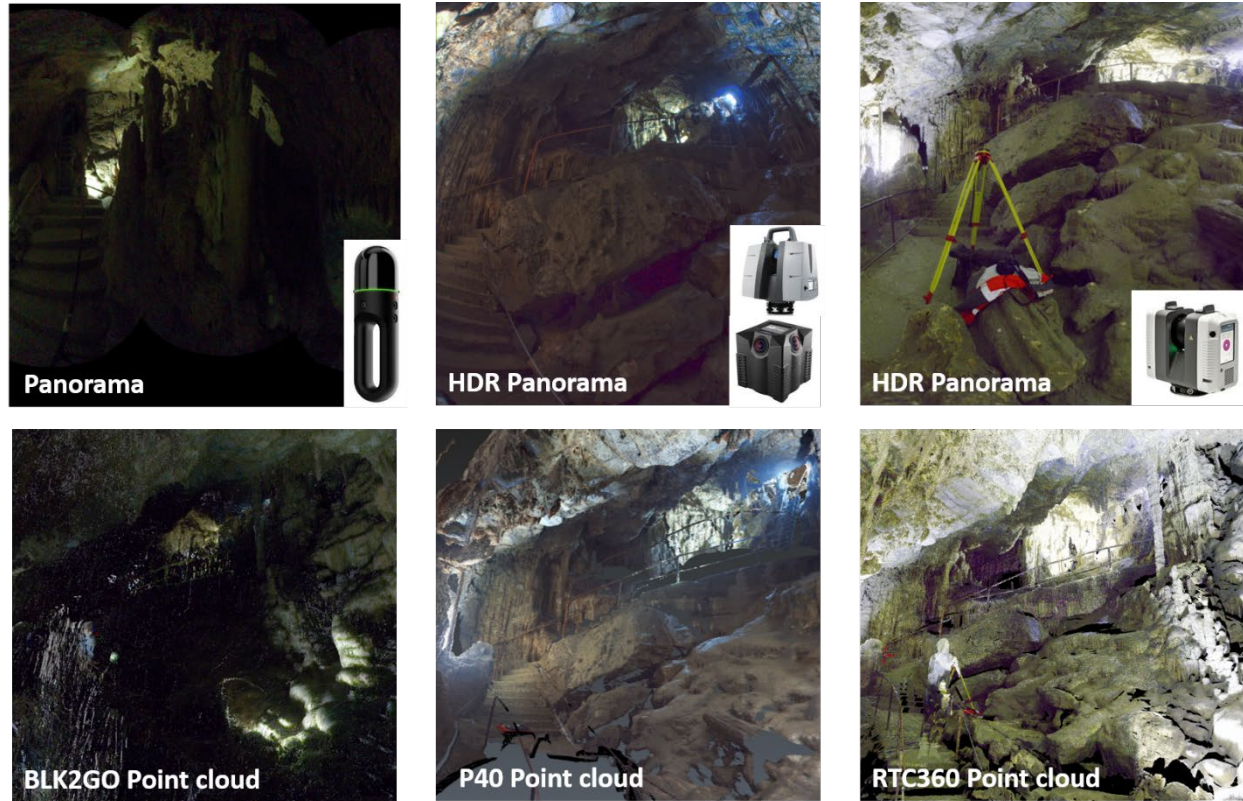


Fig. 6: Different panoramas with the P40, RTC360 and BLK2GO and the resulting textured point clouds

The RTC360 has internal HDR cameras that are automatically linked to the point cloud. The derived panorama images have only a slightly lower resolution compared to the Istar system (Istar 10000×5000 pixels and RTC: 8192×4096 pixels) but a better brightness correction, therefore single features come out better. However, the colour saturation is worse compared to the Istar system and true-colour-accuracy is lower.

Both systems delivered suitable panoramas for the texturing of the point clouds with a minimum of scattered light in the scenes. However, the brightness correction depended strongly on the setup location, therefore huge differences in the colour brightness can occur between overlapping point clouds of adjacent scans. The data quality was sufficient for the texturing of simple scenes as the plane rock surface of the entrance hall, because single panoramas cover big parts of the mesh and the image consistence in the texture is preserved. Whereas the resulting texture of the stalactite fields with the RTC was unsatisfying, due to the resolution of the scans and the inconsistent visibility and brightness of the panoramas caused by the complex scene and the obstructions by the stalactites.

In section 4 it was already mentioned that the structure-from-motion approach failed for the geometric reconstruction due to accuracy reasons. However, the quality of the RTC panoramas

was sufficient enough to enable a combined processing approach for the external SLR images and RTC360 point clouds with the panoramas. The software RealityCapturer was used to combine these two data sources with feature matching of the high resolution SLR images and the HDR panoramas (figure 7a). This way images from the scanner with known position (from the registration process of the scanner) have been introduced into the bundle adjustment as a constraint. Therefore, the camera locations of the SLR images have been forced onto the high accurate RTC point cloud, which enabled the texturing of the RTC360 mesh with the SLR images, as seen in figure 7b.

This procedure was tested only at selected sites of the cave yet. Capturing larger parts of the cave with SLR images and DJI drones for a combined processing with the RTC360 will be the aim of further surveying campaigns in the next years.

## 6 Virtual Reality application

Commercial VR viewers of point cloud data are provided by several laser scanner manufactures. Faro provides VR functionality within Faro Scene and Leica has updated its Cyclone VR viewer into the TrueView VR viewer. However, these products offer only basic fly throughs through point cloud data.

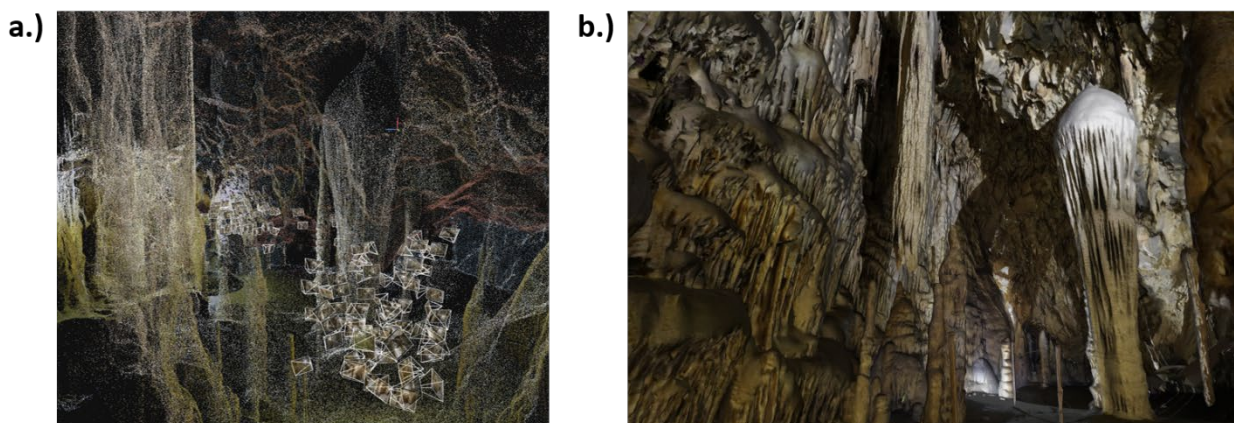


Fig. 6: a.) Screenshot from Reality Capturer with camera locations and b.) mesh with photorealistic texture of the Denkmalstein

To implement more sophisticated data interaction and visualisation options like adjusting of lighting, more immersive navigation and also basic story telling features, a custom VR application has to be made up for such occasions. Appropriate development environments are provided by game engines. These software kits were originally designed for the creation of videogames but provide a state-of-the-art 3D-object-oriented environment for numerical simulation and data interaction. Basic physic engines and light engines support the creation of realistic simulations. Due to their close relation to the video game market the implementation of VR devices into the application can be established with open source plug-ins.

Several game engines can be found on the market like Unity, Unreal or the cry engine. Although the Unreal engine supports point clouds and has advantages in photorealistic texturing the



application for the Katerloch was made up in Unity. Here, the creator can build the application from custom C# routines like in any other coding environment, which gives the environment a huge degree of freedom and a big support of the community. As an exchange format between geodetic software and Unity and other graphic software the obj-format for 3D objects has been found suitable. It is the most basic mesh format, which is human readable, has a great interoperability with software from the entertainment sector and contains also texture information as a jpg-file. It can be loaded easily into Unity. Mesh colliders can be attached and the user can physically interact with the environment in VR.

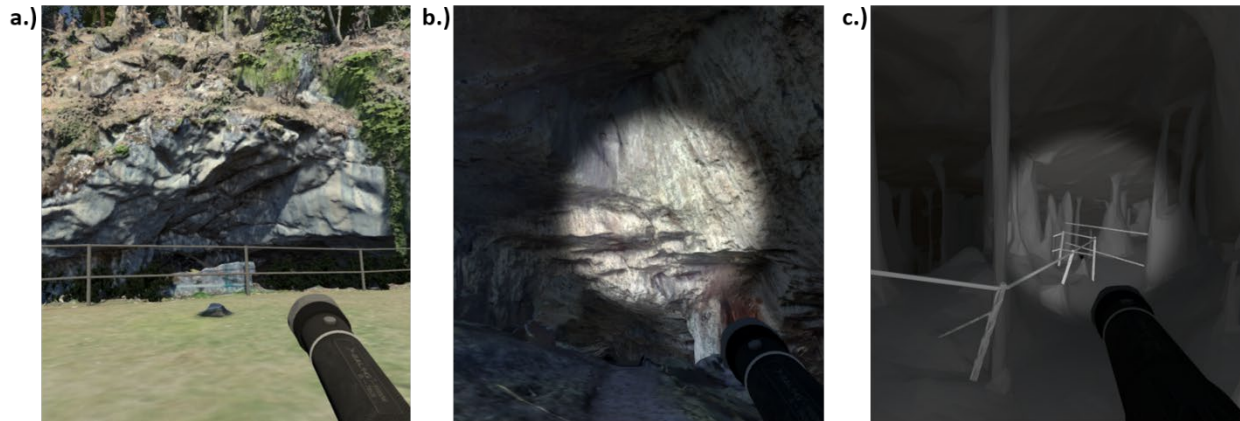


Fig. 7: a.) High resolution texture b.) Low resolution texture with virtual torch light c.) untextured part in the VR application

Based on the investigations of DANYLUK & WILLET (2019), to give the user a realistic cave experience in the application it was the aim to provide the possibility to move along the 3D model to explore and investigate every detail of the dataset. The HTC Vive Pro has an outside-in tracking system and provides all six degrees of freedom for user movement with an accuracy of several millimetres (BAUER et al. 2021). Therefore, the actual user movement is applied in 1:1 scale on the virtual counterpart and it is necessary to duck in narrow passages to get through. Furthermore, two handheld controllers with touchpads are used to interact with data sets and to apply user movement beyond the capabilities of the HTC Vive Pro tracking system.

Early versions of the simulation, especially by younger audience, were not received so well due to the missing of the texture, because the main focus was placed on presenting the mesh quality and dimensions of the cave. Different user groups have confirmed that introducing also parts with poor texture in the simulation is even better than homogeneously sparing out the texture at the entire mesh. Therefore, the different texture qualities have become an integral part of the story telling of the final VR application. The user starts at a highly textured and nearly photorealistic entrance area (derived from DJI drone imagery) to familiarize with the virtual environment and the controls, as seen in figure 8a. As the users progresses on the tourist path down the cave, there is a slow transition to lower resolution textures from the Istar camera or the RTC360 panoramas. For smoothing this transition, the lighting in the scene gets darker and the user has to rely more and more on a virtual torch (figure 8b). This ends up in nearly complete darkness (except the light spot from the torch) when the user reaches the untextured BLK meshes in the heart of the cave in figure

8c. This does not only cover up the variations in the quality, but also supports the feeling of venturing down the deep cave.

Along the user path additional information is provided as images and information panels, so the user can learn about the background of the Katerloch and the geodetic survey, as seen in figure 9a. 3D models of the used laser scanner are also placed in the scene to represent the scanner setup locations and to teach students about the optimal scanner placement and provide an immersive best practice example (figure 9b).

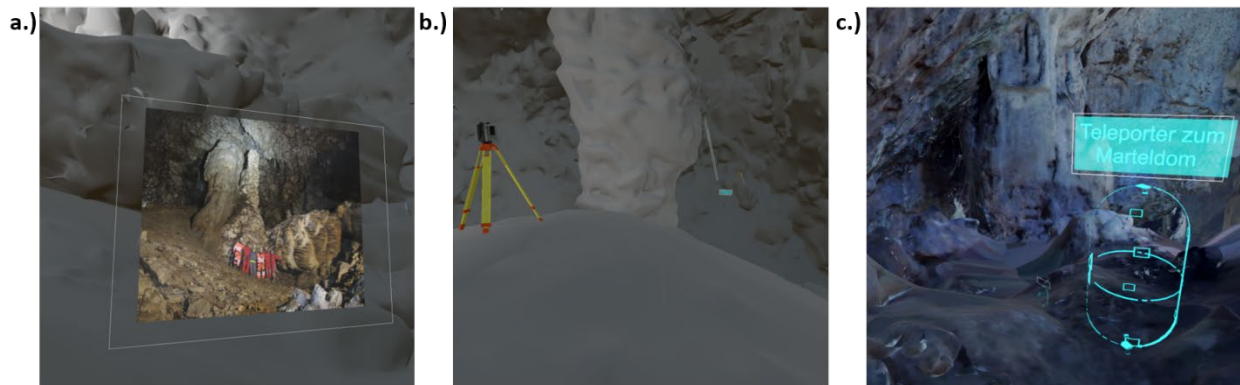


Fig. 8: a.) canvas with image b.) setup location of the RTC360 c.) teleporter for loading new scenes

Due to the mesh resolution and the amount of presented data, not all information can be loaded at once. In the real cave the visitor has to make use of ladders several times on the path. Not only is this hard to represent in VR applications, but also do these passages provide natural breakpoints in the visitor path. Therefore, the data set is separated in three levels, which are loaded sequentially and the user can change the scenes by accessing virtual teleporters, as seen in figure 9c.

## 7 Summary

Fast and mobile scanners solutions provide today the possibility of gathering a massive amount of data in a very short period of time. However, this often happens without an external reference only based on slam technology or cloud to cloud matching. This is prone for vast systematic deviations that cannot be revealed with relative quality inspection approaches. The campaigns in the Katerloch have demonstrated that several high-resolution scans with accurate registration through total station measurements provide a solid data foundation for more mobile solutions to capture challenging areas. Not only can these scans be co-registered into a global coordinate system, but the quality can be assessed by the investigation of the transformation parameters.

The meshing algorithms of point cloud data are making huge progresses, due to advances in the entertainment sector, who are also starting the use 3D surveying data for their movie set designs or for virtual worlds in video games. The geodetic software 3DReshaper provides here a straight forward automatic meshing approach, that relies strictly on the given data set. Therefore, the resulting mesh needs a lot of manual clean up to fill holes and merging parts. On the other hand, software solutions like RealityCapturer from Epic are able to produce a watertight mesh without much need of postprocessing. Within these software solutions, the algorithm takes many decisions

and automatically masks out lamps and other smaller objects. Although the resulting photorealistic mesh is good for presentations the detailed documentation of the scene is questionable up to a degree. This can be especially problematic when using such intelligent meshing approaches for sensitive documentations of crime scenes, because the resulting mesh is highly affected by many hypotheses.

The implementation of VR technology was a valid attempt for presenting complex 3D data to a diverse audience. The ego perspective of VR technologies takes over when conventional viewers reach their limits with overlapping geometries and multi temporal datasets. Nevertheless, the uncomfortable long-term use is still an issue with state-of-the-art devices. The VR application of the Katerloch is today a basic tool at public events to promote 3D scanning and the field of geodesy.

With the advances in software and hardware, especially in the combination of SLR images and scans is much potential for the creation of a photorealistic model of the entire cave, which will be the aim of future measurement campaigns.

## 8 References

- BAUER, P., LIENHART, W. & JOST, S., 2021: Accuracy Investigation of the Pose Determination of a VR System. *Sensors*, **21**, 1622. <https://doi.org/10.3390/s21051622>.
- BUCHROITHNER, M. & GAISECKER, T., 2020: Ice surface changes in Eisriesenwelt (Salzburg, Austria) based on LIDAR measurements between 2017 and 2020, *DIE HÖHLE - Zeitschrift für Karst und Höhlenkunde*, 71. Jg. Verband Österreichischer Höhlenforschung.
- CHANG, E., YOO & B., KIM, H., 2020: Virtual Reality Sickness: A Review of Causes and Measurements, Article in *International Journal of Human-Computer Interaction*, <https://doi.org/10.1080/10447318.2020.1778351>.
- DANYLUK, K. & WILLETT, W., 2019: Evaluating the Performance of Virtual Reality Navigation Techniques for Large Environments. *Advances in Computer Graphics*, [https://doi.org/10.1007/978-3-030-22514-8\\_17](https://doi.org/10.1007/978-3-030-22514-8_17).
- DETERDING, S., KHALED, R., NACKE, L. & DIXON, D., 2011: Gamification: Toward a definition. Conference: CHI 2011 Gamification Workshop Proceedings, 12-15.
- DETERDING, S., 2015: The Lens of Intrinsic Skill Atoms: A Method for Gameful Design. *Human-Computer Interaction*, **30**, 294-335, <https://doi.org/10.1080/07370024.2014.993471>.
- DE WAELE, J., FABRI, S., SANTAGATA, T., CHIARINI, V., COLUMBU, A. & PISANI, L., 2018: Geomorphological and speleogenetical observations using terrestrial laser scanning and 3D photogrammetry in a gypsum cave (Emilia Romagna, N. Italy), *Geomorphology*, **319**, <https://doi.org/10.1016/j.geomorph.2018.07.012>.
- GEISSLER, F., 2023: History of the Katerloch. <https://www.katerloch.at/index.php/die-hoehle/geschichte>, last access 30.01.2023.
- GEIST, M., NIEMAYER, F. & GIERSCHNER, F., 2018: Modellierung – Strategien zur Interpretation von Punktwolken. 177. DVW-Seminar "Terrestrisches Laserscanning 2018 (TLS 2018), Fulda, Germany, Schriftenreihe des DVW, **93**, 59-74.

- GEOFFREY, J., 2020: The early days of 3d scanning, xyht digital magazine. <https://www.xyht.com/surveying/the-early-days-of-3d-scanning-cyra/>, last access 30.01.2023.
- HOLST, C., JOST, B. & KUHLMANN, H., 2016: TLS-basierte Deformationsanalyse unter Nutzung von Standardsoftware. 154. DVW-Seminar "Terrestrisches Laserscanning 2016 (TLS 2016), Fulda, Germany, Schriftenreihe des DVW, **85**, 39-58.
- LEICA GEOSYSTEMS, 2016: Leica ScanStation P30/P40. Heerbrugg, Switzerland, 839696en – 12.20. Downloaded from: <https://leica-geosystems.com/en-us/products/laser-scanners/scanners/leica-scanstation-p40--p30>, last access 30.01.2023.
- LEICA GEOSYSTEMS, 2018: Leica RTC360 Datasheet. Heerbrugg, Switzerland 2018. 872750en - 04.22. Downloaded from: <https://leica-geosystems.com/en-us/products/laserscanners/scanners/leica-rtc360>, last access 30.01.2023).
- LEICA GEOSYSTEMS, 2020: Leica RTC360: Slam in terrestrial Laser scanning. Heerbrugg, Switzerland 917247en – 03.20. Downloaded from: <https://leica-geosystems.com/en-us/products/laserscanners/scanners/leica-rtc360>, last access 30.01.2023)
- LEICA GEOSYSTEMS, 2023a: Leica BLK360 Specification sheet. Downloaded from: <https://leica-geosystems.com/en-us/products/laser-scanners/scanners/blk360>, last access 30.01.2023.
- LEICA GEOSYSTEMS, 2023b: Leica BLK2GO Datasheet. Downloaded from: <https://leica-geosystems.com/en-us/products/laser-scanners/autonomous-reality-capture/leica-blk2go-handheld-imaging-laser-scanner>, last access 30.01.23.
- VIRTUALREALITYSOCIETY, 2017: History of Virtual Reality, Homepage accessed: <https://www.vrs.org.uk/virtual-reality/history.html>, last access 08.04.2022.
- WALTERS, R. & HAJNA, N., 2020: 3D laser scanning of the natural caves: example of Škocjanske jame. Geodetski vestnik. **64**, <https://doi.org/10.15292/geodetski-vestnik.2020.01.89-103>.

# Analyse der Echtzeit-Visualisierung für Virtual Reality-Anwendungen am Beispiel der Festung Al Zubarah in Katar

THOMAS P. KERSTEN<sup>1</sup>, DANIEL DRENKHAN<sup>1</sup> & SIMON DEGGIM<sup>1</sup>

*Zusammenfassung: Die technologischen Fortschritte der letzten Jahre im Bereich der Virtuellen Realität (VR) haben das Potenzial, unser tägliches Leben immer stärker zu beeinflussen. VR ermöglicht es, eine digitale Welt mit einem Head-Mounted Display (HMD) als immersives Erlebnis zu erkunden. In Kombination mit Werkzeugen zur 3D-Dokumentation, Modellierung und Software zur Erstellung interaktiver virtueller Welten kann VR eine wichtige Rolle bei der Erhaltung und Visualisierung des kulturellen Erbes für Museen, Bildungseinrichtungen und andere kulturelle Bereiche spielen. Entsprechende Game Engines bieten Werkzeuge zur interaktiven 3D-Visualisierung von Kulturobjekten, was eine neue Form der Wissensvermittlung durch eine direkte Beteiligung der Anwender in der virtuellen Welt ermöglicht. Um eine reibungslose Echtzeit-Visualisierung der Daten im HMD zu gewährleisten, sollten VR-Anwendungen jedoch mit 90 Bildern pro Sekunde laufen. Diese Bildrate ist von verschiedenen Kriterien wie Datenmenge oder Anzahl der dynamischen Objekte abhängig. In diesem Beitrag wurden Untersuchungen zur Leistungsfähigkeit einer VR-Anwendung anhand verschiedener digitaler 3D-Modelle mit unterschiedlichen Datenmengen des Forts Al Zubarah in Katar durchgeführt. Wir zeigen, welchen Einfluss die Datenmenge und die Hardware-Ausstattung auf die Performance der Echtzeit-Visualisierung haben und dass Entwickler von VR-Anwendungen daher einen Kompromiss zwischen Datenmenge und Leistungsfähigkeit der verfügbaren Computerhardware finden sollten, um eine flüssige Echtzeit-Visualisierung mit ca. 90 fps (frames per second) zu gewährleisten. CAD-Modelle bieten aufgrund des deutlich geringeren Datenvolumens eine bessere Leistung für die Echtzeit-VR-Visualisierung als vermaschte 3D-Modelle.*

## 1 Einführung

Virtuelle Realität (VR) ermöglicht eine neue Form der Präsentation und Visualisierung von Kulturdenkmälern. Immersive Erlebnisse in VR bieten interessante Möglichkeiten in der Denkmalpflege, Informationen und Wissen so an ein breites Publikum zu vermitteln. Diese Entwicklung schafft somit spannende neue Möglichkeiten für Institutionen, die sich mit Kulturobjekten beschäftigen, wie z.B. Museen, mit ihren Sammlungen und Archiven neue Zielgruppen anzusprechen. Insbesondere mithilfe digitaler 3D-Rekonstruktionstechniken erlaubt VR dem Publikum, historische Orte im Maßstab 1:1 in einer räumlich immersiven visuellen und auditiven Umgebung zu erleben.

Es gibt verschiedene Definitionen von VR, die das Verständnis für diese Technologie erleichtern. VR ist die Darstellung und gleichzeitige Wahrnehmung der Realität und ihrer physikalischen Eigenschaften in einer in Echtzeit computergenerierten, interaktiven virtuellen Umgebung. Darüber hinaus ist die virtuelle Realität eine realistische und immersive Simulation einer dreidimensionalen

---

<sup>1</sup> HafenCity Universität Hamburg, Labor für Photogrammetrie & Laserscanning, Henning-Voscherau-Platz 1, D-20457 Hamburg, E-Mail: [Thomas.Kersten, Simon.Deggim]@hcu-hamburg.de, Daniel.Drenkhan@web.de

Umgebung, die mit interaktiver Software und Hardware geschaffen wird und durch Bewegungen des Körpers erlebt oder gesteuert wird. So verstanden ist VR eine künstliche Umgebung, die durch sensorische Stimuli (wie perspektivische Ansichten und Geräusche) erlebt wird, die von einem Computer bereitgestellt werden, wobei eigene Handlungen (teilweise) bestimmen, was in der Umgebung geschieht (MERRIAM-WEBSTER DICTIONARY 2023). Diese sehr weit gefasste Definition erlaubt es, die meisten modernen Anwendungen von VR zu berücksichtigen. Weitere Definitionen finden sich in der Literatur von DÖRNER et al. (2014), FREINA & OTT (2015), und PORTMAN et al. (2015). LANIER (1992) beschreibt die technische Ausstattung und die Anforderungen, die notwendig sind, um die Illusion zu erzeugen, sich in einer virtuellen Welt zu befinden. Der Begriff wurde erstmals vom Autor Damien Broderick in seinem 1982 erschienenen Science-Fiction-Roman „Das Judas-Mandala“ eingeführt. Bereits 1962 entwickelte Morton Heilig das Sensorama, eine Maschine, die eines der frühesten bekannten Beispiele für immersive, multisensorische (heute als multimodal bezeichnete) Technologie darstellt und somit als das erste VR-System bezeichnet werden kann. 1968 entwickelte Ivan Sutherland mit Unterstützung seiner Studenten Bob Sproull, Quintin Foster und Danny Cohen den ersten Prototyp eines VR-Systems, das an einen Computer angeschlossen wurde und modernen VR-Systemen ähnlich ist (RHEINGOLD 1992, DUBOSE 2020). Die VR-Technologie ist somit nicht neu, aber erst mit der Einführung kostengünstiger VR-Headsets - z. B. der Oculus Rift im Jahr 2014 und der HTC Vive im Jahr 2016 - ist diese Technologie für ein breiteres Publikum anwendbar geworden. Heutige VR-Anwendungen profitieren besonders von der Echtzeit-Visualisierung inklusive Interaktionsmöglichkeiten und Sensortracking mittels Head-Mounted Displays (HMDs), was den Grad der Immersion erhöht. Eine erste empirische Überprüfung HMD vs. Bildschirmdarstellung wurde von HRUBY et al. (2020) durchgeführt. Um diese Technologie für ein immersives Erlebnis nutzen zu können, müssen die folgenden Voraussetzungen erfüllt sein: (a) eine virtuelle 3D-Umgebung muss in einer integrierten Entwicklungsumgebung (z. B. in einer Game-Engine) konstruiert und texturiert werden (können), (b) die entwickelte und ausführbare VR-Anwendung muss über eine geeignete Software (z. B. Steam VR) mit einem Head-Mounted Display (HMD) verbunden werden, und (c) die Bewegungen des Nutzers müssen über Controller und HMD gesteuert und verfolgt werden können.

Ein wichtiger Faktor in VR ist die Immersion, die den Effekt beschreibt, der durch eine VR-Umgebung erzeugt wird, die das Bewusstsein des Benutzers, illusorischen Reizen ausgesetzt zu sein, so weit in den Hintergrund treten lässt, dass die virtuelle Umgebung als real wahrgenommen wird. Um eine reibungslose, immersive Visualisierung der VR-Anwendung im VR-Headset zu gewährleisten, sollte die Anzahl der Bilder pro Sekunde (fps) jedoch idealerweise mindestens 90 betragen. Je niedriger die Bildrate der VR-Anwendung im HMD ist, desto wahrscheinlicher ist es, dass der Nutzer aufgrund der Latenzzeit leichter Übelkeit oder Schwindel (Motion Sickness oder Cyber Sickness) erfährt, was zu allgemeinem Unwohlsein führen kann, während er die VR-Brille trägt. Eine große Herausforderung bei der Entwicklung einer VR-Anwendung ist daher, eine Balance zwischen qualitativ hochwertiger Darstellung der Daten und Visualisierung dieser in Echtzeit zu finden. In diesem Beitrag wird am Beispiel der Festung Al Zubarah die Leistungsfähigkeit von VR-Anwendungen untersucht. Zu diesem Zweck wurden 3D-Punktwolken aus terrestrischem Laserscanning (TLS) und Structure-from-Motion (SFM) Photogrammetrie verwendet, um vermaschte 3D-Modelle mit unterschiedlichen Auflösungen und ein CAD-Modell (Computer-Aided

Design) zu erstellen. Wichtige Faktoren, die sich auf die Leistung der VR-Anwendung (gemessen in gerenderten Bildern pro Sekunde) auswirken können, sind u.a. die Datenmenge und die Texturierung der erzeugten Modelle sowie die unterschiedliche Hardwareausstattung. Es stellt sich also die Frage, welchen Einfluss die Datenmenge und die Hardware-Ausstattung auf die Leistung einer VR-Anwendung bei der Echtzeit-Visualisierung haben.

Im zweiten Kapitel wird ein Überblick über andere Arbeiten in diesem Themenbereich gegeben, während im dritten Kapitel die Festung Al Zubarah vorgestellt wird. Nach der Vorstellung der 3D-Modellierung im vierten Kapitel wird im fünften Kapitel die Erstellung der VR-Anwendung beschrieben. Die Ergebnisse der Untersuchungen zur Leistungsfähigkeit der VR-Anwendung werden im sechsten Kapitel vorgestellt. Abschließend wird ein Fazit aus den Untersuchungen gezogen und ein Ausblick gegeben.

## 2 Vorherige Arbeiten

VR spielt bei der Entwicklung und Darstellung des virtuellen Kulturerbes eine entscheidende Rolle. Sie eröffnet eine neue Form der öffentlichen und wissenschaftlichen Kommunikation, insbesondere für historische Kulturobjekte und Denkmäler, die entweder bereits beschädigt oder zerstört oder zu weit von potenziellen interessierten Besuchern entfernt stehen (ADDISON 2000; STONE & OJIKI 2005; AFFLECK & THOMAS 2005). POLIMERIS & CALFOGLOU (2016) versuchen, die Wirkung des digitalen Mediums Virtual Reality näher zu beleuchten, indem sie in einer kleinen Studie die Auswirkungen verschiedener Präsentationsformen kulturtouristischer Produkte auf die Entscheidung der Befragten für ein kulturtouristisches Reiseziel vergleichen. MEDYŃSKA-GULIJ & ZAGATA (2020) bewerteten in einer Fallstudie anhand der Festung in Ostrów Lednicki (Polen) die Wirkung des Eintauchens in einen spezifischen historisch-geografischen virtuellen Raum für Experten und Spieler. BOZORGI & LISCHER-KATZ (2020) stellen das Virtual Ganjali Khan Project vor, eine laufende Forschungsinitiative, die 3D- und VR-Technologien einsetzt, um die Erhaltung der kulturellen Stätte des Ganjali Khan Gebäudeensembles zu unterstützen, einer großen historischen Sehenswürdigkeit in der Wüstenstadt Kerman, Iran. EDLER et al. (2019) stellen vor, wie VR-basierte 3D-Umgebungen (basierend auf der Spiel-Engine Unreal Engine 4) zur Unterstützung der Stadtteilentwicklung eines umstrukturierten postindustriellen Areals anhand des VR-Modells "Zechen Holland" in Bochum-Wattenscheid als repräsentatives ehemaliges Industriegebiet im deutschen Ruhrgebiet eingesetzt werden können.

An der HafenCity Universität Hamburg wurden bereits mehrere VR-Anwendungen für kulturelle Stätten und Objekte entwickelt. Das Museum in Bad Segeberg, das in einem Fachwerkhaus aus dem 16. Jahrhundert untergebracht ist, wurde digital in sechs Bauphasen (von 1541 bis 1963) als VR-Erlebnis mit der HTC Vive Pro konstruiert (KERSTEN et al. 2017). Drei historische Städte in Deutschland (sowie deren Umgebung) wurden als VR-Erlebnisse entwickelt: Duisburg im Jahr 1566 (TSCHIRSCHWITZ et al. 2019), Segeberg im Jahr 1600 (DEGGIM et al. 2017, KERSTEN et al. 2018), und Stade im Jahr 1620 (WALMSLEY & KERSTEN 2019). Darüber hinaus sind auch drei religiöse und kulturelle Denkmäler als VR-Erlebnisse verfügbar: die Selimiye-Moschee in Edirne, Türkei (KERSTEN et al. 2017), ein Holzmodell des Salomonischen Tempels (KERSTEN et al. 2018) und der Kaiserdom in Königslutter bei Braunschweig mit 360°-Panoramafotos in eine immersive

Echtzeit-Visualisierung integriert (WALMSLEY & KERSTEN 2020). Ein weiteres Beispiel für eine immersive und interaktive VR-Präsentation eines Monuments sind die İnceğiz-Höhlen im Istanbuler Stadtteil Çatalca, Türkei, die mithilfe von Punktwolken aus terrestrischen Laserscans in 3D modelliert und in die Unity 3D Game Engine integriert wurden (BÜYÜKSALIH et al. 2020).

Die Zahl der Arbeiten, die sich speziell mit der Echtzeit-VR-Visualisierung von Kulturdenkmälern befassen, wächst. Zu den jüngsten Museumsexponaten, die Echtzeit-VR zur Visualisierung des kulturellen Erbes nutzen, gehören u.a. auch Batavia 1627 im Westfries Museum in Hoorn, Niederlande (WESTFRIES MUSEUM 2023), und Viking VR, das als Begleitung einer Ausstellung im Britischen Museum entwickelt wurde (SCHOFIELD et al. 2018). Eine Reihe neuerer Forschungsprojekte befasst sich ebenfalls mit dem Einsatz von VR zur Visualisierung des kultureller Denkmäler (FASSI et al. 2016; SEE et al. 2016; SKARLATOS et al. 2016; RAMSEY 2017; DHANDA et al. 2019) sowie mit Aspekten, die über die Visualisierung hinausgehen, einschließlich der Nachbildung physischer Umweltreize (MANGHISI et al. 2016). Diese Literaturliste ist jedoch nicht vollständig.

Veröffentlichungen zu Studien über die Leistung von Echtzeit-VR-Visualisierungen sind eher begrenzt. KHARROUBI et al. (2019) haben den Einfluss der Größe der Punktwolke auf die VR-Leistung (in fps) unter Verwendung der Unity-Spielengine analysiert. Sie testeten ihren Ansatz an mehreren Datensätzen, darunter eine Punktwolke mit 2,3 Milliarden Punkten, die als Testobjekt das Schloss von Jehay (Belgien) darstellt. Ihre Ergebnisse unterstreichen die Effizienz und Leistungsfähigkeit ihrer Lösung zur Visualisierung klassifizierter massiver Punktwolken in virtuellen Umgebungen mit mehr als 100 Bildern pro Sekunde.

Die Darstellung digitaler Elemente in Virtual-Reality-Systemen erfordert sehr niedrige Latenzzeiten, um ein reibungsloses VR-Erlebnis frei von Motion Sickness und Übelkeit zu ermöglichen. Um die idealen Bedingungen für VR zu schaffen, wird eine "Motion-to-Photon-Zeit" (Zeit zwischen Sensorerkennung der Bewegung und der Reaktion auf dem Bildschirm) von weniger als 20 Millisekunden (ms), entsprechend 50 FPS (Frames per Second), angestrebt (MCCAFFREY 2017). Um die Leistung des Computers für die Echtzeit-VR-Anwendung zu untersuchen, wurden drei Szenen der Viermastbark Peking (ein historisches Segelschiff) aufgrund der unterschiedlichen Komplexität der Szenen als Beispiele ausgewählt (KERSTEN et al. 2020): Beim Rendering einer Hafensicht mit nur wenigen Umgebungsrequisiten und geringem Detailgrad wurden Spitzenwerte von bis zu 54 fps (19 ms) für die VR-Visualisierung erreicht. Mit zunehmender Anzahl von Modellen und Texturen sinkt die Leistung entsprechend, wie bei der Ansicht des gesamten Schiffes ( $\emptyset$ : 21 fps/48 ms). Durch die Simulation physikalischer Vorgänge (z. B. Wind in den Segeln) steigt die benötigte Rechenleistung stark an und die Visualisierung verzögert sich, so dass der Nutzer bei der Betrachtung der Segel ( $\emptyset$ : 15 fps/67 ms) die Latenzen aufgrund der dynamischen Vorgänge in der Szenerie im Nahbereich deutlich wahrnehmen kann. Daher wurden die Auswirkungen verschiedener Versionen der VR-Anwendung der Festung Al Zubarah auf die Leistungsfähigkeit näher untersucht, um die Auflösungsgrenzen von Modellen zu ermitteln, die für Echtzeitanwendungen wie VR momentan geeignet sind.



### 3 Die Festung Al Zubarah in Katar

Um die Leistungsfähigkeit einer VR-Anwendung unter verschiedenen Datenmengen zu untersuchen, wurde die historische Festung Al Zubarah (Abb. 1) verwendet. Die Festung befindet sich in der Gemeinde Madinat ash Shamal an der nordwestlichen Küste der Halbinsel von Katar. Mit einer Fläche von 34 m × 34 m und einer Höhe von 9 m ist diese historische katarische Militäranlage eine der bekanntesten Sehenswürdigkeiten und Touristenattraktionen in Katar. Die Festung wurde ursprünglich 1938 von Scheich Abdullah bin Jassim Al Thani erbaut und diente als Küstenwache (WIKIPEDIA 2023). Wegen des ständigen Konflikts mit dem Nachbarstaat Bahrain erhielt die Festung aufgrund der geografischen Lage eine strategische Bedeutung. Die Festung besteht aus hohen, kompakten, einen Meter dicken Mauern aus Korallen- und Kalkstein. Ein schützendes Dach aus gepresstem Lehm spendet Schatten und eine kühle Umgebung für die damaligen Soldaten des Forts, das in drei Ecken massive Rundtürme mit verschiedenen Arten von Verteidigungsanlagen enthält. Die vierte Ecke wurde als prägnanter rechteckiger Turm mit exquisiten dreieckigen Gesimsen mit Schlitzen gebaut. Im Erdgeschoss wurden acht Räume für die Unterbringung der Soldaten eingerichtet. Im Inneren des Forts befinden sich Treppen, über die man die Galerie und die Türme des Forts erreichen kann. Im Jahr 1987 wurde es zu einem Museum umgebaut, in dem verschiedene Exponate und Kunstwerke ausgestellt werden, insbesondere zeitgenössische, aktuelle archäologische Funde aus der nahe gelegenen archäologischen Ausgrabung in der ehemaligen Stadt der Perlenfischer. Von dieser historischen Stadt Al Zubarah wurden vier bedeutende Gebäude (Moschee, Palast, Türme und Markt) von FERWATI & EL MENSRAWY (2021) virtuell rekonstruiert. Seit 2013 ist diese archäologische Stätte von Al Zubarah inklusive des Forts in die Liste des UNESCO-Welterbes aufgenommen worden (THUESEN & KINZEL 2011; UNESCO 2013; UNESCO 2021).



Abb. 1: Die arabische Festung Al Zubarah in Katar - die Lage auf der Karte (links), das Außenansicht (Mitte) und der Innenbereich mit terrestrischem Laserscanner (rechts)

### 4 3D-Modellierung

Die Festung Al Zubarah wurde im September 2011 durch terrestrisches Laserscanning mit einem IMAGER 5006h von Zoller + Fröhlich (Z+F) aus Wangen im Allgäu und SFM-Photogrammetrie als Teil des Qatar Islamic Archaeology and Heritage Project für die Qatar Museums Authority erfasst und dokumentiert (KERSTEN et al. 2015). Ziel war es, das Fort in 2D und 3D zu dokumentieren, um Pläne und Schnitte für die anstehende Renovierung zu erstellen und ein 3D-Modell zu

generieren. Die digitale Konstruktion und 3D-Modellierung der Festung Al Zubarah erfolgte anhand von 3D-Punktwolken, die sowohl durch Laserscanning als auch durch Photogrammetrie generiert wurden. Folgende Varianten von 3D-Modellen wurden erstellt: (1) ein vermaschtes 3D-Modell in drei verschiedenen Auflösungen (1, 5 und 10 Millionen Dreiecke, Abb. 2 links) auf der Grundlage von terrestrischen Laserscanning-Daten, (2) ein vermaschtes 3D-Modell, das aus der Photogrammetrie unter Verwendung von 395 Fotos einer Nikon D70 (Brennweite=35 mm) aus derselben Kampagne gewonnen wurde (Abb. 2 rechts) und (3) ein optimiertes CAD-Modell, das aus der Laserscanning-Punktwolke mit der höchsten Auflösung generiert wurde (Abb. 3 und 4).

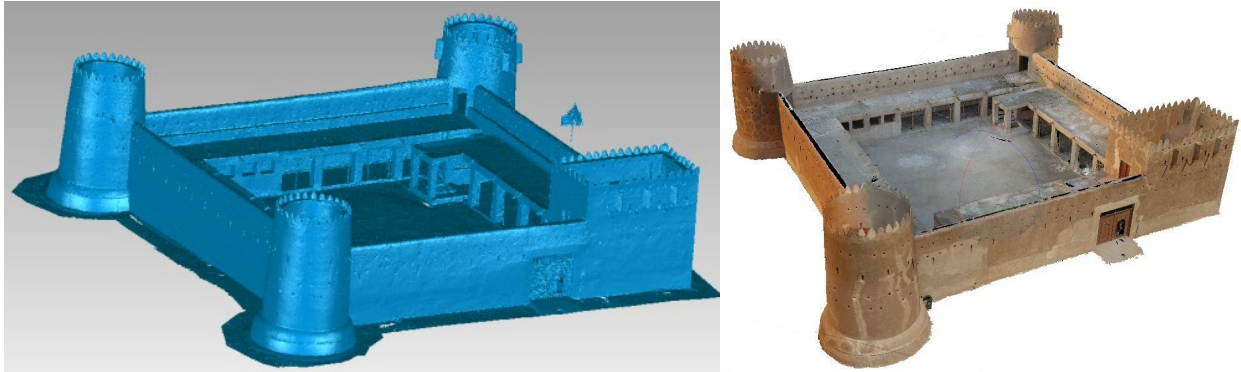


Abb. 2: 3D-Modelle der Festung Al Zubarah - Vermaschung mit 10 Millionen Dreiecken aus terrestrischem Laserscanning (links) und texturiertes 3D-Modell aus SFM-Photogrammetrie (structure-from-motion) mit 10 Millionen Dreiecken (rechts)

Die Erstellung einer Dreiecksvermaschung aus der TLS-Punktwolke wurde mit der Software Geomagic durchgeführt. Aus der Punktwolke wurden die Punkte, die die Festung repräsentieren, ausgeschnitten und alle Ausreißer eliminiert. Außerdem wurden alle Fenster und Türen ausgeschnitten, um später bewegliche Objekte wie sich öffnende Türen und Fenster in der VR-Anwendung zu berücksichtigen. Aus der Punktwolke wurde eine geschlossene Dreiecksvermaschung gerechnet, in der vorhandene Löcher manuell mit entsprechenden Funktionen in Geomagic gefüllt wurden. Von dieser ursprünglichen Vermaschung wurden drei Varianten mit verschiedenen Auflösungen von 1, 5 und 10 Millionen Dreiecken abgeleitet (Abb. 2 links). Als zweite Variante wurde aus den Fotos mit der Photogrammetrie-Software Agisoft Metashape eine Punktwolke erzeugt, aus der nach Eliminierung von Ausreißern ein RGB-farbige Dreiecksvermaschung mit 10 Millionen Dreiecken berechnet wurde. Die dritte Variante ist ein mit der Software AutoCAD konstruiertes CAD-Modell. Grundlage für die Konstruktion des CAD-Volumenmodells war die TLS-Punktwolke (Abb. 3 links). Für die Konstruktion des Grundrisses wurde zunächst in Geomagic ein horizontaler Querschnitt in der Punktwolke berechnet, der dann in AutoCAD importiert und als 2D-Linien digitalisiert wurde (Abb. 3 Mitte). Die Messungen für die Konstruktion der Festung in AutoCAD wurden in der Software Z+F LaserControl durchgeführt, in der einige der Funktionen Ebenen und Ecken erkennen können (Abb. 3 rechts). Für die Konstruktion der Festung war jedoch ein gewisses Maß an Generalisierung notwendig, da die Festung nicht immer exakt rechteckig ist und die Wände nicht immer ganz gerade sind oder die gleiche Dicke aufweisen.

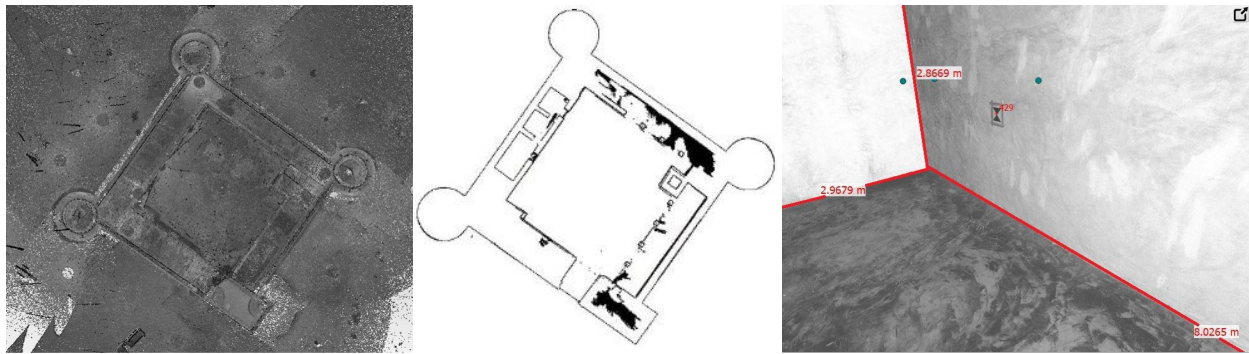


Abb. 3: TLS-Punktwolke der Festung Al Zubarah (links), der abgeleitete Grundriss (Mitte) und Messungen in der Punktwolke (rechts)

## 5 Entwicklung der VR-Applikation

Für die Entwicklung der VR-Anwendung wurden die in Kapitel 4 beschriebenen Modelle des Forts als fbx-Dateien in die Game Engine Unreal Engine Version 4.24 importiert. Eine Game Engine ist eine Simulationsumgebung, in der 2D- oder 3D-Grafiken durch entsprechenden Programmcode manipuliert werden können. Sie wurden in erster Linie für die Videospieleindustrie entwickelt, bieten aber auch ideale Plattformen für die Erstellung von VR-Erlebnissen, da viele der erforderlichen Funktionen bereits implementiert sind. Für dieses Projekt wurde die Unreal Engine gewählt, da sie den Vorteil einer implementierten visuellen Programmierung durch sogenannte Blueprints bietet, die den Anwendern ermöglichen, einfache Interaktionen und Animationen ohne Vorkenntnisse in der Programmiersprache C++ zu entwickeln. Der allgemeine Arbeitsablauf für die Erstellung und Visualisierung einer VR-Anwendung ist in Abb. 4 schematisch dargestellt. Allerdings wurden die 3D-Modelle nicht wie üblich in der 3DS-Software texturiert, sondern zunächst in der Game-Engine. Lediglich das vermaschte 3D-Modell aus den photogrammetrischen Daten wurde bereits in der Software Metashape texturiert.

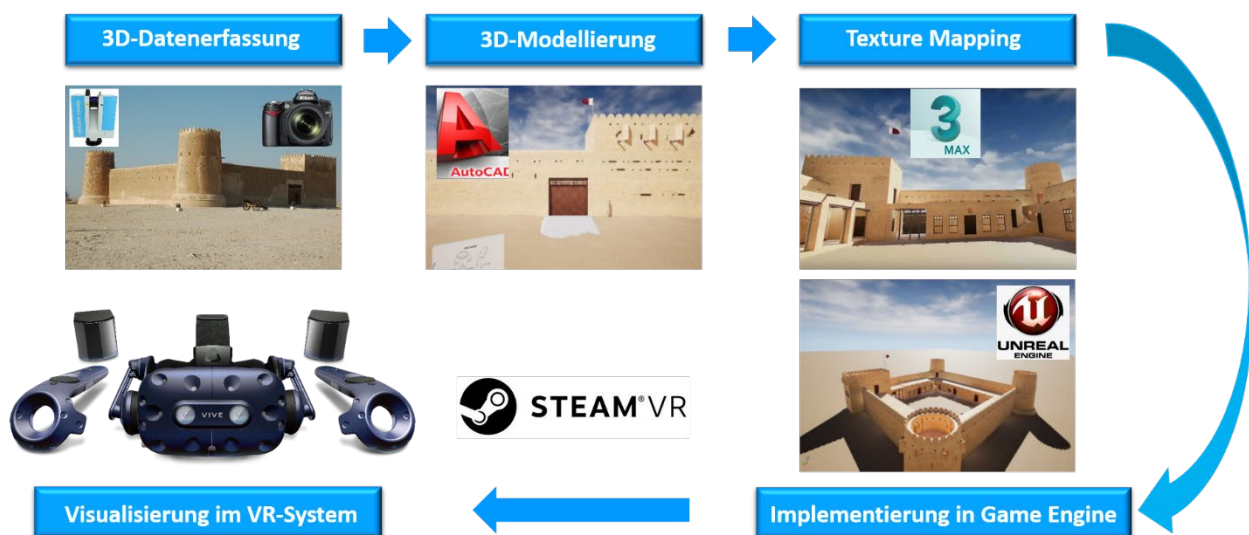


Abb. 4: Workflow für die Entwicklung der VR-Anwendung der Festung Al Zubarah

Das Gebiet um das Fort Al Zubarah besteht aus einer flachen Wüstenlandschaft. Daher wurde die Landschaft um das Fort in der Spiel-Engine als eine einfache flache Ebene konstruiert. Die Umgebung wurde mit einigen Büschen und Bäumen in einer Ecke des Forts ausgestattet. Die Bewegung und Navigation des Benutzers innerhalb der VR-Anwendung erfolgt durch Teleportation (Abb. 5 links und Mitte), die mit den Controllern gesteuert wird. Um die Bewegung des Benutzers in der Anwendung einzuschränken, wurde eine unsichtbare Grenze um das Fort gezogen, die es dem Benutzer erlaubt, sich nur innerhalb des Forts und in der unmittelbaren Umgebung zu bewegen. Um zu verhindern, dass sich der Benutzer durch die Festungsmauer teleportiert, wurde für jedes Element ein komplexes Kollisionsnetz erstellt. Bei beweglichen Objekten, wie z. B. Türen, ändert sich das Kollisionsnetz, so dass sich der Benutzer durch offene Türen teleportieren kann (Abb. 5 Mitte und rechts). Die Türen öffnen sich automatisch, wenn sich der Benutzer innerhalb einer Triggerbox bewegt, die sich vor und hinter der Tür befindet. Neben der eigenen Bewegung innerhalb der Anwendung hat der Benutzer mehrere Möglichkeiten, über die beiden programmierten Controller mit der Anwendung zu interagieren, z. B. das Schalten verschiedener Lichtquellen, die im Fort zum Betrieb entsprechender Lampen in den Innenräumen und über den Bildern in der Ausstellung genutzt werden können. Diese Interaktion wird durch das Bewegen des Controllers zum Lichtschalter ausgelöst. Eine weitere Interaktion innerhalb der VR-Anwendung ist die per Triggerbox ausgelöste Teleportation des Nutzers in jedem Turm über eine Leiter auf die darüber liegende Ebene.

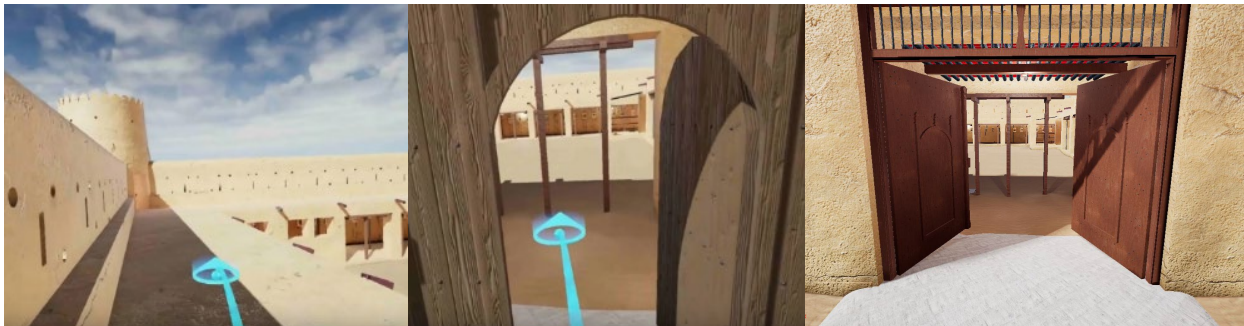


Abb. 5. Teleportation des Nutzers in der VR-Anwendung (links und Mitte), das Öffnen der Tür als Interaktion (rechts)



Abb. 6. Perspektivische Ansichten der entwickelten VR-Anwendung anhand des CAD-Modells der Festung: Mauer mit rechteckigem und rundem Turm (links) und Innenhof (rechts)

Die endgültige VR-Anwendung (Abb. 6) ermöglicht es dem Benutzer, in Echtzeit zwischen den Versionen zu wechseln und sowohl den kulturellen Aspekt - die digitale Version der Festung - als

auch den technologischen Aspekt mit den Auswirkungen verschiedener Formen der Nutzung von 3D-Datensätzen und deren Arbeitsabläufen zur Erstellung digitaler Modelle zu erkunden.

## 6 VR-Performance-Tests und Ergebnisse

Um aussagekräftige Ergebnisse für die Leistungstests zu erhalten, wurden für die Leistungsanalyse fünf verschiedene Ansichten ausgewählt, die unterschiedliche Inhalte repräsentieren und darstellen (u. a. ein Überblick über die Festung, eine einfache Ansicht der Mauer, eine Ansicht mit wehender Flagge und eine detaillierte Innenansicht) (Abb. 7). Die Ansichten wurden aufgrund ihrer unterschiedlichen Menge an Materialien und Polygonen ausgewählt. Wir gehen davon aus, dass die unterschiedlichen Inhalte der einzelnen Ansichten einen Einfluss auf die Leistung der Echtzeit-VR-Visualisierung haben, d. h. je mehr unterschiedliche Materialien, Texturen und Polygone, desto größeren Einfluss hat es auf die VR-Leistung.

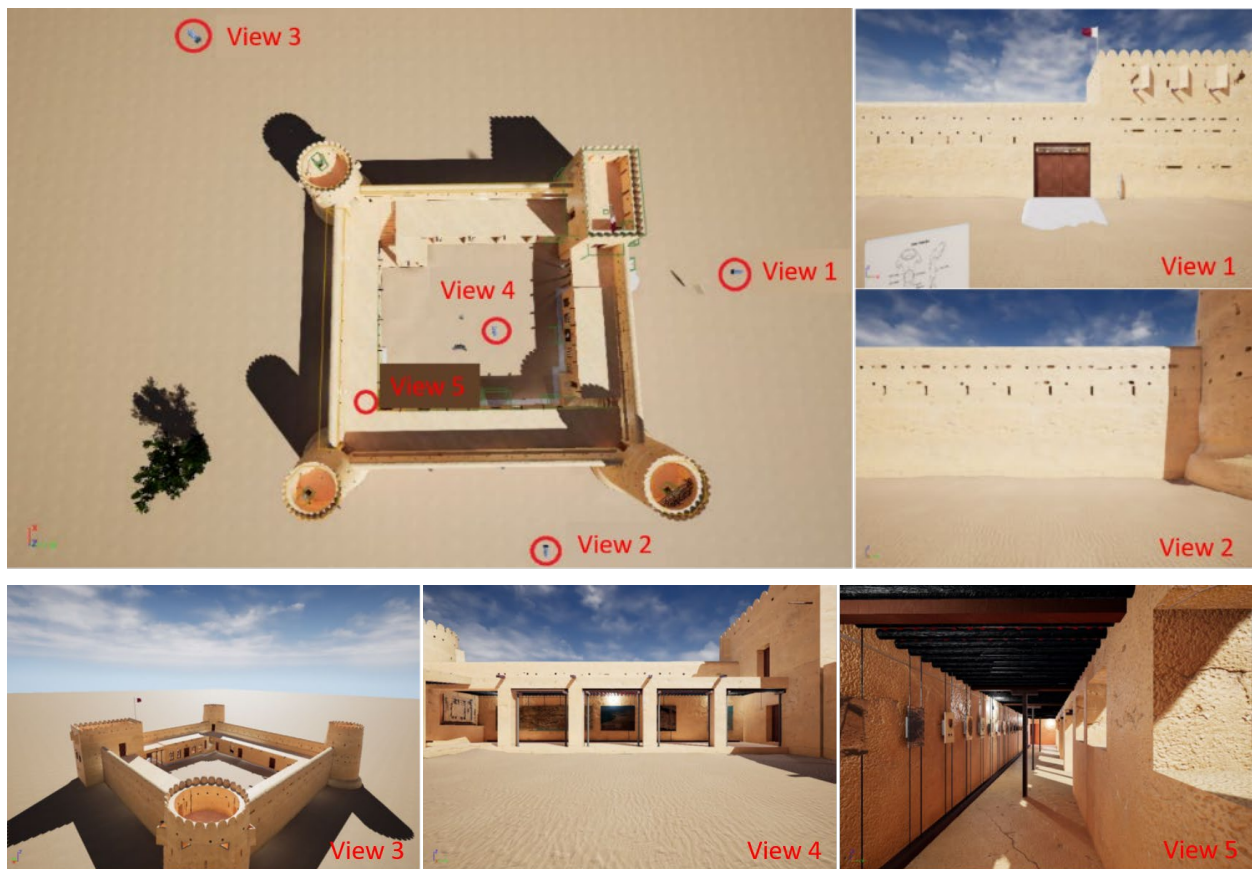


Abb. 7: Übersicht über fünf verschiedene Testszenen einschließlich ihrer perspektivischen Ansichten

Die erste Ansicht zeigt die Vorderseite des Forts mit dem einzigen Eingang und der katarischen Flagge, die auf dem rechteckigen Turm weht. Für die zweite Ansicht wurde ein Ausschnitt gewählt, in dem keine Veränderungen sichtbar sind und nur Boden, Himmel und Mauerwerk zu sehen sind. Um das gesamte Ausmaß des 3D-Modells zu erfassen, wurde eine weitere Ansicht (Ansicht 3) erstellt, die das gesamte Modell des Forts aus einem schrägen Winkel zeigt. Die vierte

Ansicht wurde in der Mitte des Innenhofs mit Blick auf die Ausstellung platziert, während die fünfte Ansicht eine Reihe von Gemälden für die Ausstellung auf der linken Seite innerhalb des schmalen Korridors zeigt.

Um den Einfluss der Hardware auf die VR-Leistung zu testen, wurden zwei VR-Anwendungen (CAD-Modell vs. vermaschtes Modell aus photogrammetrischen Bilddaten) auf drei verschiedenen Rechnern ausgeführt. Für alle anderen Tests wurde jedoch hauptsächlich der Rechner Labor 1 (Tabelle 1) verwendet, der an eine VR-Station im Labor der Hafencity Universität Hamburg angeschlossen ist und dem Nutzer ermöglicht, die Anwendung in einer virtuellen Umgebung zu erleben. Die folgenden technischen Spezifikationen der drei Computerhardware sind in Tabelle 1 zusammengefasst: CPU (Central Processing Unit), GPU (Graphics Processing Unit), und Random-Access Memory (RAM). Jede verwendete Nvidia-Grafikkarte hat eine Graphics Double Data Rate (GDDR) von acht Gigabyte.

Tab. 1: Technische Spezifikation der verfügbaren Computerhardware für die VR-Leistungstests

Computer	CPU	GPU	RAM
Labor 1	Intel Core i7 4 GHz	Nvidia GTX 1080	16 GB
Labor 2	Intel Core i7 3,4 GHz	Nvidia GTX 1080	32 GB
Home	Intel Core i5	Nvidia RTX 2060 Super	16 GB

Die Ergebnisse der VR-Leistungstests mit drei verschiedenen Computerhardware unter Verwendung einer Anwendung mit einem CAD-Modell und einem vermaschten Modell aus der SFM-Photogrammetrie sind in Abb. 8 dargestellt (hier stellen die Punkte die Messwerte dar). Aufgrund der geringeren Datenmenge des CAD-Modells im Vergleich zu den vermaschten Modellen erreichen alle drei Computer mehr als 100 fps für die VR-Anwendung bei der Verwendung der fünf verschiedenen perspektivischen Ansichten, was mehr als ausreichend ist, um die VR-Anwendung ruckelfrei auszuführen. Die leichten Unterschiede in der Leistung könnten auf die unterschiedlichen CPUs zurückzuführen sein. Bei der deutlich höheren Datenmenge der photogrammetrischen Dreiecksvermaschung zahlt sich jedoch die High-End-Grafikkarte (Nvidia RTX 2060 Super) des Home-Rechners aus, da hier immer noch über 100 fps erreicht werden. Dagegen reduziert die große Datenmenge (10 Millionen Dreiecke) die Leistung der beiden Laborrechner (Labor 1 und Labor 2), was sich in der Anzeige von jeweils nur 50 fps bei den vier Ansichten zeigt. Die Ansicht 5 war in dem aus der SFM-Photogrammetrie generierten vermaschten Modell nicht verfügbar.

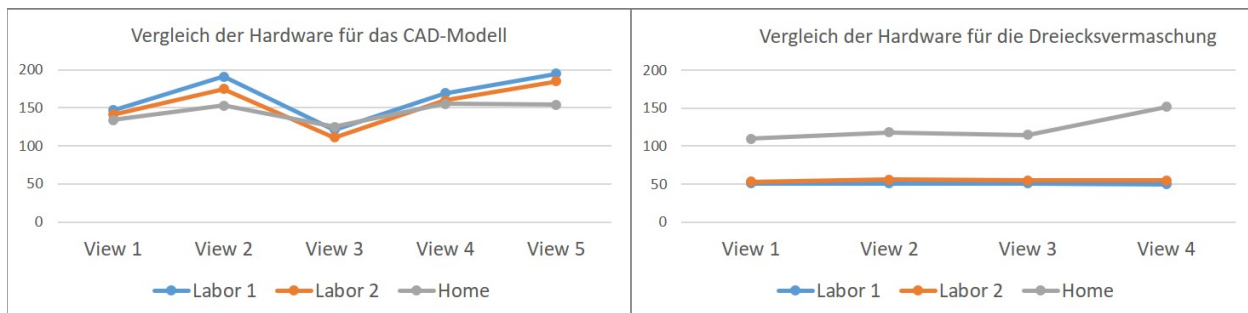


Abb. 8: Leistung der Hardware als Frames pro Sekunde (fps) in der VR-Anwendung für die verschiedenen 3D-Modelle

Der zweite Teil der Tests bestand aus einem Leistungsvergleich der verschiedenen VR-Modelle unter Verwendung derselben Computerhardware (Labor 1). Um den Einfluss anderer Faktoren so weit wie möglich auszuschließen, wurden die fps-Messungen ohne Texturen und vor der Berechnung der Beleuchtung der Szenen durchgeführt. Die Ergebnisse dieser Leistungstests sind in Abb. 9 zusammengefasst (Punkte= Messwerte). Die VR-Anwendungen mit dem CAD-Modell und dem TLS-Vermaschung mit 1 Million Dreiecken laufen mit mehr als 120 fps, während die anderen vermaschten Modelle deutlich schlechtere Leistungen bringen. Allerdings gibt es zwei interessante Aspekte in Abb. 9: (1) die TLS-Vermaschung (Mesh) mit 1 Million Dreiecken erzielt eine höhere fps-Rate als das CAD-Modell und (2) das photogrammetrisch vermaschte Modell (10 Millionen Dreiecke) von Metashape schneidet etwas besser ab als das TLS-Modell mit 10 Millionen Dreiecken. Die Gründe für beide Ergebnisse sind nicht klar. Außerdem verringert die Komplexität von Ansicht 3 die Leistung der VR-Anwendung auch bei den beiden Datensätzen mit der geringen Datenmenge (CAD-Modell und TLS 1M Mesh).

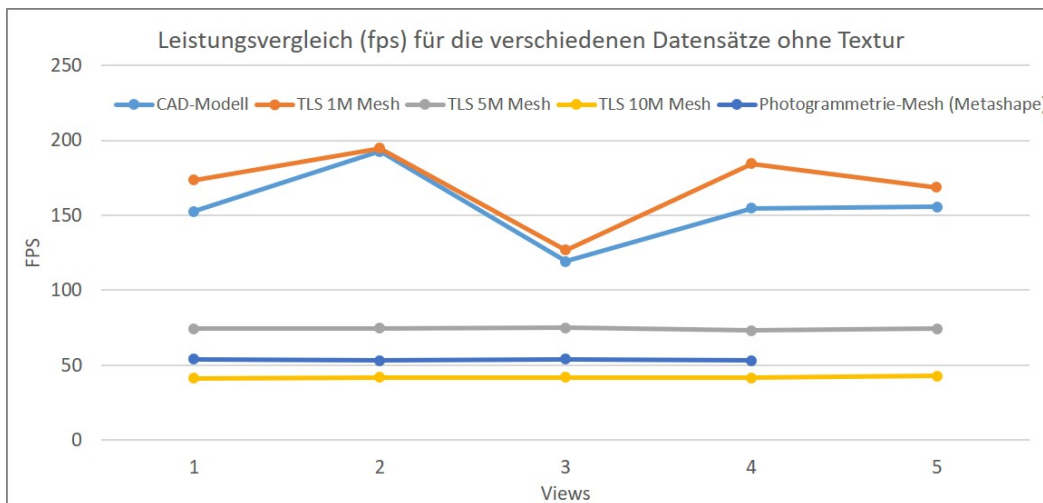


Abb. 9: Leistungsvergleich (in fps) der verschiedenen Datensätze ohne Textur (für Ansichten 1-5)

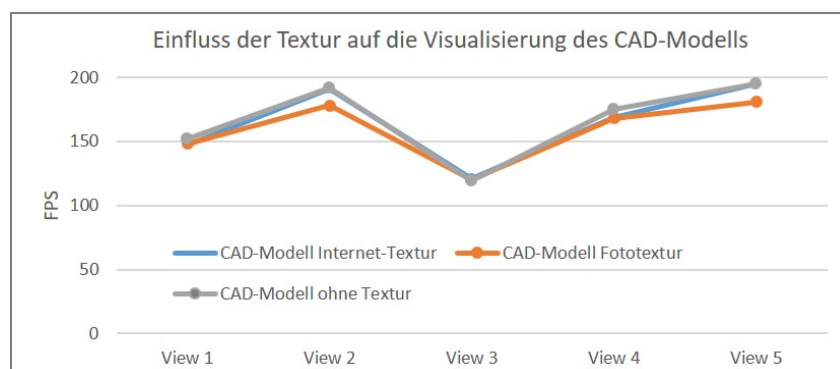


Abb. 10: Einfluss der Textur auf die Computerleistung bei der Visualisierung des CAD-Modells für die Ansichten 1-5

Aufgrund der hohen Computerleistung hat die Textur hier nur sehr geringen Einfluss auf die Visualisierung des CAD-Modells, da die Bildraten für alle fünf Ansichten jeweils deutlich über 100 fps liegen (Abb. 10) und da die Differenzen zwischen der Verwendung von Fototextur und keiner

Textur bei maximal 14 fps. Dagegen zeigt Abb. 11 den Einfluss der Datenmenge in Form der vermaschten TLS-Modelle mit Variation in Millionen von Dreiecken auf die fps-Rate für die beiden komplexeren Ansichten 1 und 3. Für diese Analyse wurde die Datenmenge in der VR-Anwendung künstlich erhöht, um den Effekt des Datenwachstums in den Ansichten 1 und 3 deutlich zu visualisieren. So konnte gezeigt werden, dass es derzeit eine Begrenzung der Datenmenge von ca. 5 Millionen Dreiecken in der Game-Engine Unreal 4 gibt, um eine reibungslose Echtzeit-Visualisierung der VR-Anwendung mit der verwendeten Computerhardware und den entsprechenden Einstellungen zu gewährleisten.

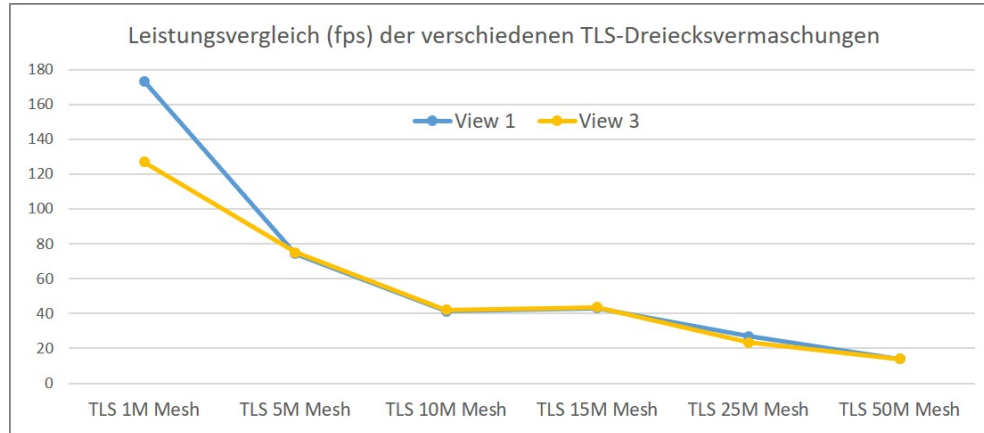


Abb. 11: Frames pro Sekunde (fps) als Messpunkte für die verschiedenen TLS-Modelle (Meshes), die in VR-Anwendungen verwendet werden

## 7 Fazit und Ausblick

In diesem Beitrag wird die Entwicklung einer VR-Anwendung der katarischen Festung Al Zubarah in der Game Engine Unreal 4 als Grundlage für eine Performance-Analyse vorgestellt. Damit wurde die Leistungsfähigkeit der VR-Anwendung anhand verschiedener Kriterien untersucht, wobei die Leistung von ca. 90 Bildern pro Sekunde für eine optimale Echtzeit-Visualisierung im HMD vorausgesetzt wurde. Für diesen Test wurden fünf verschiedene 3D-Modelle generiert und in eine VR-Anwendung integriert: drei vermaschte 3D-Modelle in verschiedenen Auflösungen (1, 5 und 10 Millionen Dreiecke), die auf terrestrischen Laserscanning-Daten basieren, ein vermaschtes 3D-Modell, das aus Daten der Photogrammetrie gewonnen wurde, und ein optimiertes CAD-Modell, das aus der Laserscanning-Punktwolke konstruiert wurde. Bei den Leistungstests wurden folgende Aspekte untersucht: Leistung der Computerhardware und die fünf verschiedenen 3D-Modelle mit unterschiedlicher Dreiecksanzahl. Die Leistungsuntersuchungen in der Spiele-Engine Unreal Engine 4 zeigten, dass es immer wichtig ist, einen Kompromiss zwischen der Datenmenge und der verfügbaren Computerhardware zu finden, um eine flüssige Echtzeit-Visualisierung mit mehr als 90 fps zu gewährleisten. Daher ist es von Vorteil, CAD-Modelle anstelle von vermaschten Modellen zu verwenden, da die Datenmenge signifikant geringer ist. Bei der Verwendung eines vermaschten Modells liegt die Obergrenze für eine optimierte Echtzeit-VR-Visualisierung mit der verwendeten Hard- und Software sowie den Einstellungen jedoch bei 5 Millionen Dreiecken. Je



höher die Auflösung der vermaschten Modelle ist, desto geringer ist die Leistung der VR-Anwendungen in Form von Geschwindigkeit (Bilder pro Sekunde). Diesem Effekt kann durch eine Hardware-Ausstattung mit leistungsfähiger Grafikkarte und höherem Computerspeicher (Random-Access-Memory) entgegengewirkt werden. Des Weiteren kann festgestellt werden, dass die unterschiedlichen Inhalte der einzelnen generierten perspektivischen Ansichten der Festung einen erheblichen Einfluss auf die Performance der VR- Echtzeit-Visualisierung haben.

Darüber hinaus gibt es verschiedene andere Elemente, die die Leistung einer VR-Anwendung stark beeinflussen können, z. B. die Qualität der generierten und der dynamischen Schatten, die Unterteilung großer 3D-Modelle in viele kleinere, die Detailgenauigkeit (Level of Detail LOD), die Lichteinstellungen, die Umgebungseinstellungen, die Rendering-Pipeline der gewählten Spiel-Engine usw. Dieser Beitrag konzentriert sich jedoch auf die Unreal Engine und auf den Vergleich mit aus Punktwolkendaten abgeleiteten vermaschten Modelle (Meshes). Die übliche vermaschte Punktwolke ist ein einziges großes Mesh, was für Echtzeitanwendungen ungewöhnlich ist und eigene Herausforderungen hinsichtlich der Leistung mit sich bringt, z. B. das große Datenvolumen einer hochauflösenden vermaschten Punktwolke. Optimierungsmethoden für eine ähnliche Anwendung werden z. B. von LÜTJENS et al. (2019) vorgestellt, die zeigen konnten, dass große, hochauflösende Geländedatensätze in einer Virtual-Reality-Anwendung durch die Einbeziehung von Kacheln, Level-Streaming und LOD-Algorithmen visualisiert werden können. Für die Zukunft ist zu erwarten, dass die Leistung der Computerhardware und der Game-Engines deutlich zunehmen wird, so dass auch die flüssige Darstellung größerer Datenmengen in VR möglich wird.

## 8 Literaturverzeichnis

- ADDISON, A.C., 2000: Emerging Trends in Virtual Heritage. *IEEE MultiMedia*, 7(2), 22-25.
- AFFLECK, J. & THOMAS, K., 2005: Reinterpreting Virtual Heritage. 10th International Conference on Computer Aided Architectural Design Research in Asia, 1, 169-78.
- BOZORGI, K. & LISCHER-KATZ, Z., 2020: Using 3D/VR for Research and Cultural Heritage Preservation: Project Update on the Virtual Ganjali Khan Project. *Preservation, Digital Technology & Culture*, 49(2), 45-57, <https://doi.org/10.1515/pdte-2020-0017>.
- BÜYÜKSALIH, G., KAN, T., ÖZKAN, G. E., MERIC, M., ISYIN, L. & KERSTEN, T., 2020: Preserving the Knowledge of the Past Through Virtual Visits: From 3D Laser Scanning to Virtual Reality Visualisation at the Istanbul Çatalca İnceğiz Caves. *Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 88, 133-146, <https://doi.org/10.1007/s41064-020-00091-3>.
- DEGGIM, S., KERSTEN, T., TSCHIRSCHWITZ, F. & HINRICHSSEN, N., 2017: Segeberg 1600 - Reconstructing a Historic Town for Virtual Reality Visualisation as an Immersive Experience. *Inter. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, 42(2/W8), 87-94, <https://doi.org/10.5194/isprs-archives-XLII-2-W8-87-2017>.
- DHANDA, A., REINA ORTIZ, M., WEIGERT, A., PALADINI, A., MIN, A., GYI, M., SU, S., FAI, S. & SANTANA QUINTERO, M., 2019: Recreating cultural heritage environments for VR using photogrammetry. *Int. Arch. Photogrammetry, Remote Sens. Spat. Inf. Sci.*, 42(2/W9), 305-310, <https://doi.org/10.5194/isprs-archives-XLII-2-W9-305-2019>.

- DÖRNER, R., BROLL, W., GRIMM, P. & JUNG, B., 2014: Virtual und Augmented Reality (VR/AR): Grundlagen und Methoden der Virtuellen und Augmentierten Realität; Springer-Verlag: Berlin.
- DUBOSE, J., 2020: The Case for VR. *Journal of Electronic Resources Librarianship*, **32**(2), 130-133, <https://doi.org/10.1080/1941126X.2020.1739851>.
- EDLER, D., KEIL, J., WIEDENLÜBBERT, T., SOSSNA, M., KÜHNE, O. & DICKMANN, F., 2019: Immersive VR Experience of Redeveloped Post-industrial Sites: The Example of “Zeche Holland” in Bochum-Wattenscheid. *KN J. Cartogr. Geogr. Inf.*, **69**, 267-284, <https://doi.org/10.1007/s42489-019-00030-2>.
- FASSI, F., MANDELLI, A., TERUGGI, S., RECHICHI, F., FIORILLO, F. & ACHILLE, C., 2016: VR for Cultural Heritage. *International Conference on Augmented Reality, Virtual Reality and Computer Graphics 2016, Lecture Notes in Computer Science*, **9769**, 139-157, Springer, Cham. [https://doi.org/10.1007/978-3-319-40651-0\\_12](https://doi.org/10.1007/978-3-319-40651-0_12).
- FERWATI, M.S. & EL MENSRAWY, S., 2021: Virtual reconstruction of the historic city of Al-Zubarah in Qatar. *Digital Applications in Archaeology and Cultural Heritage*, **21**, e00177, <https://doi.org/10.1016/j.daach.2021.e00177>.
- FREINA, L. & OTT, M., 2015: A Literature Review on Immersive Virtual Reality in Education: State of The Art and Perspectives; *eLearning & Software for Education*. Available online: <https://ppm.itd.cnr.it/download/eLSE%202015%20Freina%20Ott%20Paper.pdf>, last access 11 January 2023.
- HRUBY, F., SÁNCHEZ, L.F.Á., RESSL, R. & ESCOBAR-BRIONES, E.G., 2020: An Empirical Study on Spatial Presence in Immersive Geo-Environments. *Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, **88**, 155-163, <https://doi.org/10.1007/s41064-020-00107-y>.
- KERSTEN, T., BÜYÜKSALIH, G., TSCHIRSCHWITZ, F., KAN, T., DEGGIM, S., KAYA, Y. & BASKARACA, A. P., 2017: The Selimiye Mosque of Edirne, Turkey - An Immersive and Interactive Virtual Reality Experience using HTC Vive. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, **42**(5/W1), 403-409, <https://doi.org/10.5194/isprs-archives-XLII-5-W1-403-2017>.
- KERSTEN, T., DEGGIM, S., TSCHIRSCHWITZ, F., LINDSTAEDT, M. & HINRICHSSEN, N., 2018: Segeberg 1600 - Eine Stadtrekonstruktion in Virtual Reality. *KN J Cartogr Geogr Inf* **68**(4), 183-191, <https://doi.org/10.1007/BF03545360>.
- KERSTEN, T., MECHELKE, K. & MAZIULL, L., 2015: 3D Model of Al Zubarah Fortress in Qatar - Terrestrial Laser Scanning vs. Dense Image Matching. *Inter. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, **40**(5/W4), 1-8, <https://doi.org/10.5194/isprsarchives-XL-5-W4-1-2015>.
- KERSTEN, T., TRAU, D. & TSCHIRSCHWITZ, F., 2020: The Four-masted Barque Peking in Virtual Reality as a New Form of Knowledge Transfer. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, **V-4-2020**, 155-162, <https://doi.org/10.5194/isprs-annals-V-4-2020-155-2020>.
- KERSTEN, T., TSCHIRSCHWITZ, F., DEGGIM, S., 2017: Development of a Virtual Museum including a 4D Presentation of Building History in Virtual Reality. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **42**(2/W3), 361-367, <https://doi.org/10.5194/isprs-archives-XLII-2-W3-361-2017>.

- KERSTEN, T., TSCHIRSCHWITZ, F., LINDSTAEDT, M. & DEGGIM, S., 2018: The historic wooden model of Solomon's Temple: 3D recording, modelling and immersive virtual reality visualisation. *Journal of Cultural Heritage Management and Sustainable Development, Special Issue: Digital workflows for the conservation and sustainability of historic places*, **8**(4), 448-464, <https://doi.org/10.1108/JCHMSD-09-2017-0067>.
- KHARROUBI, A., HAJJI, R., BILLEN, R. & POUX, F., 2019: Classification and Integration of Massive 3D Points Clouds in a Virtual Reality (VR) Environment. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, **42**(2/W17), 165-171, <https://doi.org/10.5194/isprs-archives-XLII-2-W17-165-2019>.
- LANIER, J., 1992: Virtual reality: The promise of the future. *Interactive Learning International*, **8**(4), 275-279.
- LÜTJENS, M., KERSTEN, T., DORSCHER, B. & TSCHIRSCHWITZ, F., 2019: Virtual Reality in Cartography: Immersive 3D Visualization of the Arctic Clyde Inlet (Canada) Using Digital Elevation Models and Bathymetric Data. *Multimodal Technol. Interact.*, **3**(9), <https://doi.org/10.3390/mti3010009>.
- MANGHISI, V.M., FIORENTINO, M., GATTULLO, M., BOCCACCIO, A., BEVILACQUA, V., CASCELLA, G.L., DASSISTI, M. & UVA, A.E., 2018: Experiencing the sights, smells, sounds, and climate of southern Italy in VR. *IEEE Computer Graphics and Applications 2018*, **37**(6), 19-25, <https://doi.org/10.1109/MCG.2017.4031064>.
- MCCAFFREY, M., 2017: *Unreal Engine VR Cookbook: Developing Virtual Reality with UE4*. Addison-Wesley Professional, 1st edition, 288 p.
- MEDYŃSKA-GULIJ, B. & ZAGATA, K., 2020: Experts and Gamers on Immersion into Reconstructed Strongholds. *ISPRS International Journal of Geo-Information* **9**(11), 655, <https://doi.org/10.3390/ijgi9110655>.
- MERRIAM-WEBSTER DICTIONARY, 2023: Definition of virtual reality, <https://www.merriam-webster.com/dictionary/virtual%20reality>, last access 11.01.2023.
- POLIMERIS, S. & CALFOGLOU, C., 2016: Cultural Tourism Destinations and the Power of Virtual Reality. In: Katsoni V., Stratigea A. (eds) *Tourism and Culture in the Age of Innovation*. Springer Proceedings in Business and Economics. Springer, Cham, [https://doi.org/10.1007/978-3-319-27528-4\\_39](https://doi.org/10.1007/978-3-319-27528-4_39).
- PORTMAN, M.E., NATAPOV, A. & FISHER-GEWIRTZMAN, D., 2015: To go where no man has gone before: Virtual reality in architecture, landscape architecture and environmental planning. *Comput. Environ. Urban Syst.*, **54**, 376-384.
- RAMSEY, E., 2017: Virtual Wolverhampton: Recreating the historic city in virtual reality. *ArchNet International Journal of Architectural Research* **2017**, **11**(3), 42-57, <http://dx.doi.org/10.26687/archnet-ijar.v11i3.1395>.
- RHEINGOLD, H., 1992: *Virtual Reality: The Revolutionary Technology of Computer-Generated Artificial Worlds - and How It Promises to Transform Society*. New York: Simon & Schuster
- SCHOFIELD, G., BEALE, G., BEALE, N., FELL, M., HADLEY, D., HOOK, J., MURPHY, D., RICHARDS, J. & THRESH, L., 2018: Viking VR: Designing a Virtual Reality Experience for a Museum. *DIS 2018 – Proceedings of the Designing Interactive Systems Conference, ACM DIS Conference on Designing Interactive Systems 2018*, Association for Computing Machinery (ACM), 805-816, <https://doi.org/10.1145/3196709.3196714>.

- SEE, Z.S., SANTANO, D., SANSOM, M., FONG, C.H. & THWAITES, H., 2018: Tomb of a Sultan: A VR Digital Heritage Approach. 3rd Digital Heritage International Congress (Digital HERITAGE) held jointly with 24th IEEE International Conference on Virtual Systems & Multimedia (VSMM 2018), 1-4.
- SKARLATOS, D., AGRAFIOTIS, P., BALOGH, T., BRUNO, F., CASTRO, F., PETRIAGGI, B.D., DEMESTICHA, S., DOULAMIS, A., DRAP, P., GEORGOPOULOS, A., KIKILLOS, F., KYRIAKIDIS, P., LIAROKAPIS, F., POUILLIS, C. & RIZVIC, S., 2016: Project iMARECULTURE: advanced VR, iMmersive serious games and augmented REality as tools to raise awareness and access to European underwater CULTURAl heritagE. Euro-Mediterranean Conference 2016, Springer, Cham, 805-813.
- STONE, R. & OJKA, T., 2000: Virtual Heritage: what next? IEEE MultiMedia, **7**(2), 73-74.
- THUESEN, I. & KINZEL, M., 2011: Al-Zubarah Archaeological Park as a UNESCO World Cultural Heritage Site - a master plan for its site management, preservation, and presentation (poster). Proceedings of the Seminar for Arabian Studies, **41**, 371-376.
- TSCHIRSCHWITZ, F., RICHERZHAGEN, C., PRZYBILLA, H.-J. & KERSTEN, T., 2019: Duisburg 1566: Transferring a Historic 3D City Model from Google Earth into a Virtual Reality Application. PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science, **87**(1-2), 47-56, <https://doi.org/10.1007/s41064-019-00065-0>.
- UNESCO, 2013: Qatar and Fiji get their first World Heritage sites as World Heritage Committee makes six additions to UNESCO List. <https://whc.unesco.org/en/news/1045>, last access 11.01.2023.
- UNESCO, 2021: Al Zubarah Archaeological Site. <https://whc.unesco.org/en/list/1402>, last access 11.01.2023.
- WALMSLEY, A. & KERSTEN, T., 2019: Low-cost development of an interactive, immersive virtual reality experience of the historic city model Stade 1620. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci., **42**, 405-411, <https://doi.org/10.5194/isprs-archives-XLII-2-W17-405-2019>.
- WALMSLEY, A. & KERSTEN, T., 2020: The Imperial Cathedral in Königslutter (Germany) as an Immersive Experience in Virtual Reality with Integrated 360° Panoramic Photography. MDPI Journal Applied Sciences, Special Issue Augmented Reality, Virtual Reality & Semantic 3D Reconstruction, **10**, 1517; <https://doi.org/10.3390/app10041517>.
- WESTFRIES MUSEUM, 2023: Batavia 1627 in virtual reality. Hoorn, Netherlands. <https://wfm.nl/batavia-1627vr>, last access 11.01.2023.
- WIKIPEDIA, 2023: Al Zubara Fort. [https://en.wikipedia.org/wiki/Al\\_Zubara\\_Fort](https://en.wikipedia.org/wiki/Al_Zubara_Fort), last access 11 January 2023.

# AUTORENVERZEICHNIS

## A

AASEN, HELGE ..... 197

## B

BADECKI, TOMASZ ..... 251

BARTELT, PERRY ..... 276

BASCHEK, BJÖRN ..... 263

BAUMANN, CHRISTOPH ..... 30

BAUER, PETER ..... 368

BECK, CLEMENS ..... 358

BENDIG, JULIANE ..... 150

BERGER, RALF ..... 110

BLEIER, MICHAEL ..... 332

BOGNER, MARKUS ..... 251

BORN, JÜRGEN ..... 301

BRÄUER-BURCHARDT, CHRISTIAN ..... 332

BRANDMEIER, MELANIE ..... 142

BRECHTKEN, RAINER ..... 233

BRIECHLE, SEBASTIAN ..... 213

BRUNN, ANSGAR ..... 98, 251

BUCHER, TILMAN ..... 110

BUDDE, LINA E. .... 301

BÜHLER, YVES ..... 276

## C

COLLMAR, DAVID ..... 163

## D

DAMBERG, SIMON ..... 233

DASSER, GWENDOLYN ..... 276

DEDRING, NIELS ..... 61

DETTMER, CHRISTIAN ..... 332

DEGGIM, SIMON ..... 381

DRENKHAN, DANIEL ..... 381

DUCKSTEIN, ALEXANDRA ..... 30

## E

EICHHORN, ANDREAS ..... 310

## F

FROMME, DIRK ..... 332

**G**

GESSLER, BASTIAN .....	263
GESSNER, MATTHIAS .....	110
GOEBEL, MONA .....	87
GÖLZ, KATRIN .....	288
GONSCHOREK, JULIA .....	110
GRAF, LUKAS V. ....	197
GRAW, VALERIE .....	61
GREIWE, ANSGAR .....	233
GREZA, MICHAEL .....	50, 174
GRÖMER, MICHAEL .....	322
GÜLCH, EBERHARD .....	36

**H**

HÄNDLER, VERENA .....	251
HEIN, DANIEL .....	110
HELL, MAXIMILIAN .....	142
HEINZE, MATTHIAS .....	332
HEIPKE, CHRISTIAN .....	223
HIRT, PHILIPP-ROMAN .....	174
HNATUSHENKO, VIKTORIIA .....	223
HNATUSHENKO, VOLODYMYR .....	223
HOCHSTUHL, SYLVIA .....	288
HOEGNER, LUDWIG .....	174
HOU, JIWEI .....	87
HÜTT, CHRISTOPH .....	206

**I**

IWASZCZUK, DOROTA .....	87, 301, 310
-------------------------	--------------

**J**

JEDMOWSKI, CHRISTOPH .....	150
----------------------------	-----

**K**

KASHTAN, VITA .....	223
KE, BINGXIN .....	69
KERSTEN, THOMAS P. ....	123, 381
KIANI SHAHVANDI, MOSTAFA .....	30
KÖLLE, MICHAEL .....	163
KUCHENDORF, CHRISTINA .....	150
KRAFT, THOMAS .....	110
KRÜGER, NINA .....	134

**L**

LAUTERBACH, HELGE ANDREAS .....	332
LENZ, NINO .....	50
LINDSTAEDT, MAREN .....	123
LIENHART, WERNER .....	368

**M**

MAIWALD, FERDINAND .....	358
MANCONI, ANDREA .....	276
MANDLBURGER, GOTTFRIED .....	1, 183, 322
MARX, STEFAN .....	332
MECHERNICH, SILKE .....	263
MEIßNER, HENRY .....	110
MUNKELT, CHRISTOPH .....	332
MEYER, JONAS .....	22
MINNICH, THEA .....	98
MÜNSTER, SANDER .....	358
MUNCH, JESSICA .....	276

**N**

NEUNER, HANS .....	183
NÜBEL, HANNES .....	78
NÜCHTER, ANDREAS .....	142, 332

**O**

OLBERDING, HELGE .....	347
------------------------	-----

**P**

PFEIFER, NORBERT .....	183
PLABST, STEPHAN .....	134
PÖPPL, FLORIAN .....	183

**R**

RASCHER, UWE .....	150
REUß, OLIVER .....	251
PFENNIGBAUER, MARTIN .....	183
RIENOW, ANDREAS .....	61
ROSCHLAUB, ROBERT .....	174
ROTH, JONAS .....	251

**S**

SAHIB, LAYTH .....	301
SALATTNA, SAJA .....	150
SCHINDLER, KONRAD .....	69
SCHIRMER, UWE .....	251
SCHLIENKAMP, ANDREAS .....	233
SCHMIDT, JAKOB .....	310
SCHNELLE, MIKE .....	123
SELENSKI, JÜRGEN .....	251
SOJA, BENEDIKT .....	30
SÖRGEL, UWE .....	163
SPRECKELS, VOLKER .....	233
STILLA, UWE .....	174
STUCKER, CORINNE .....	69

**T**

THIELE, ANTJE .....	288
TUTZAUER, PATRICK .....	78

**U**

ULLRICH, ANDREAS .....	183
------------------------	-----

**W**

WAGNER, ANDREAS .....	251
WALMSLEY, ALEXANDER P. ....	123
WALTER, VOLKER .....	163
WENG, ADRIAN .....	36
WESTERHAUS, MALTE .....	30
WESTFELD, PATRICK .....	1
WILD, FRANZISKA .....	36
WILDHABER, SAMUEL .....	197



# PROGRAMMKOMITEE

**Abmayr, Thomas**  
**Bareth, Georg**  
**Bongartz, Jens**  
**Brunn, Ansgar**  
**Coors, Volker**  
**Cramer, Michael**  
**Gerke, Markus**  
**Gülch, Eberhard**  
**Haala, Norbert**  
**Hastedt, Heidi**  
**Hochschild, Volker**  
**Hoegner, Ludwig**  
**Jacobsen, Karsten**  
**Jung, András**  
**Kaufmann, Viktor**  
**Kersten, Thomas P.**  
**Kleinschmit, Birgit**  
**Krzystek, Peter**  
**Luhmann, Thomas**  
**Mandlbürger, Gottfried**  
**Mocnik, Franz-Benjamin**  
**Motagh, Mahdi**  
**Nebiker, Stephan**  
**Reigber, Andreas**  
**Reiterer, Alexander**  
**Schmitt, Michael**  
**Schneider, Danilo**  
**Sörgel, Uwe**  
**Spreckels, Volker**  
**Vohland, Michael**  
**Wagenknecht, Stefan**  
**Weinmann, Martin**  
**Westerholt, René**

