

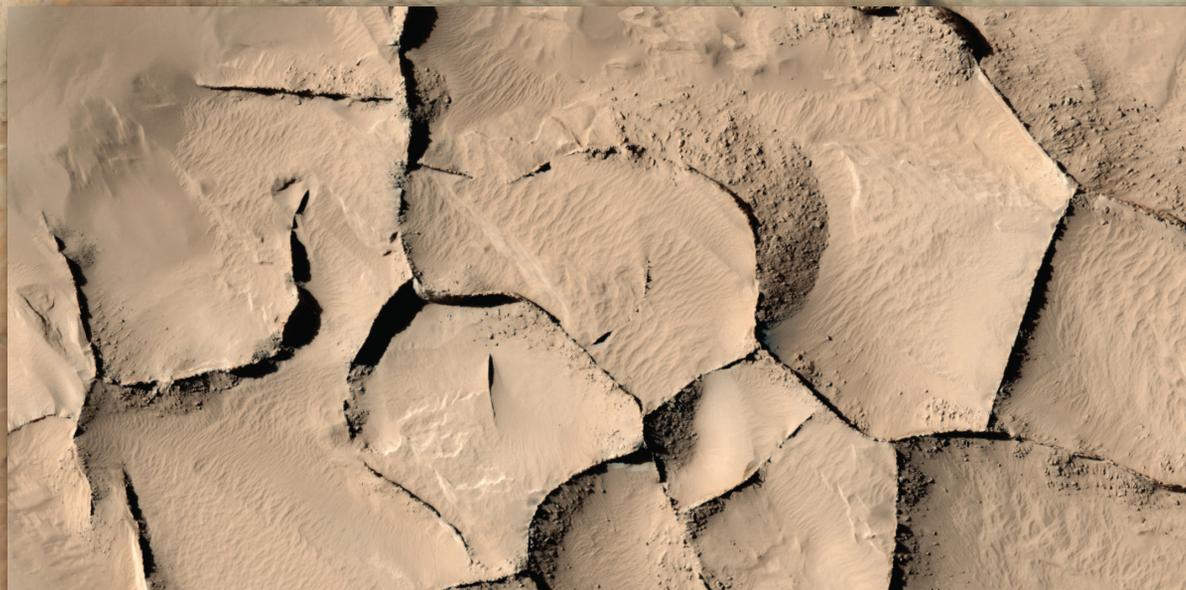
PE&RS

June 2019

Volume 85, Number 6

Shiprock

Intrusive Dike



The official journal for imaging and geospatial information science and technology

PHOTOGRAMMETRIC ENGINEERING & REMOTE SENSING



NEW
AWARDS UP TO
\$500K



INNOVATE

TRANSFORM OUR FUTURE

R&D FUNDING PROGRAM

The National Reconnaissance Office Director's Innovation Initiative (DII) Program funds cutting-edge scientific research in a high-risk, high-payoff environment to discover innovative concepts and creative ideas that transform overhead intelligence capabilities and systems for future national security intelligence needs. The program seeks the brightest minds and breakthrough technologies from industry, academia, national laboratories, and U.S. government agencies.

Visit the website for Broad Agency Announcement and Government Sources Sought Announcement requirements.

703.808.2769



www.nro.gov/About-the-NRO/Business-Opportunities

PECORA 21 • ISRSE 38

*Continuous Monitoring of Our
Changing Planet:
From Sensors to Decisions*



Save the Date

October 6-11, 2019

Baltimore, MD • Marriott Waterfront

<https://www.asprs.org/event/pecora21-isrse38>



ANNOUNCEMENTS

Esri, the global leader in location intelligence, recognized **Dewberry** with a Partner Conference Award for Maximizing ArcGIS in Service Offerings at this year's Partner Conference. The firm was selected based on its work with Esri to integrate elevation and hydrography tools in its new ArcGIS Pro software.

For more than 30 years, Dewberry has been an industry leader in geospatial and remote sensing technologies, and recently published the third edition of the DEM Users Manual, which serves as a guide to 3D elevation technologies, products, and applications.

"Our long-standing partnership with Esri has been instrumental in the success of our geospatial practice at Dewberry," says Dewberry Associate Sid Pandey, CSM. "As geospatial technology continues to evolve at a rapid rate, we're eager to help the communities where we live and work be successful and resilient by using the data and technology that we're developing with the help of organizations like Esri."

"Having spent more than 30 years in this industry, I am thrilled to see the advances that technology is making, and we're proud to be at the forefront of those changes," states Dewberry Executive Vice President Phil Thiel.

To learn more, visit www.dewberry.com. To learn more about the DEM Users Manual, visit page 412.

Spectral Evolution's PSR+ spectroradiometer was used in an advanced environmental science class at Colorado College in Colorado Springs during January/February 2019. This intensive 3.5-week undergraduate course called "Climate Change in the High Alpine" with Dr. Ulyana Horodyskyj as the instructor, integrated science and technology, providing students the opportunity to get their hands dirty through field experiences in the high alpine regions in Colorado. It was composed of 27 students, mostly junior and senior undergraduates.

"Class assignments included completing a series of "alpine challenges" and writing up the results in technical reports," said Dr. Horodyskyj. "As an example, for one challenge students were members of a US Geological Survey team, tasked with solving a scientific problem while provided with a small budget," she added. Students were taught how to build their "field toolboxes" and how to effectively use tools such as Google Earth, Excel and SNICAR (a numerical modeling program that shows the impacts of pollution on snow), backcountry snow kits, handheld Kestrels (weather stations), aerial and underwater drones, and a full-resolution visible/near-infrared spectroradiometer sponsored by Spectral Evolution. "The course provided multiple opportunities for hands-on experience using scientific instrumentation to answer specific questions, depending on the alpine region we were studying," Dr. Horodyskyj said. "In many cases, the students did not have prior experience with the instrumentation, including the spectroradiometer, and they commented on the ease of use, especially in challenging field conditions."

In order to determine how pollutants such as black carbon and natural contaminants (like dust) impacted the snowpack, the

PSR+ was useful in providing snow "spectral fingerprints." The dirtier the snow, the less the reflectance, which is clearly visible in the spectrum. The less reflectance, the more the snow can absorb solar radiation and melt faster. Spectral changes, given amounts of snow contaminants, can be numerically modeled using the online SNICAR tool mentioned above and compared with actual field data collected from the PSR+. In order to determine whether such drops in reflectance were due simply to natural causes (such as snow grain size differences), students also collected field data with backcountry snow kits. "Given its high resolution, ease of use, and clean data (we collected data with a lens and not a contact probe), the PSR+ was incredibly useful in the field for snow/ice applications," Horodyskyj said.

For more information on SPECTRAL EVOLUTION, visit www.spectralevolution.com.

GeoCue Group announced today the release of Get3Di, a geospatial data portal specifically designed to make finding, accessing, and downloading our nation's vast collection of public Lidar and Image data sources (3Di) as easy as any online shopping experience.

Get3Di differs from other online geospatial data portals by offering public data that can be used for any purpose, royalty free. Subscription prices are set to recover data download costs.

Hank DiPietro, GeoCue's VP of Business Development, says, "Our mission with Get3Di is to make public 3D data easy to find and download. Currently, to get quality public data you must look through multiple sites or try to download the data from a crowded server. We designed Get3Di to solve that problem. Our goal is to make Get3Di the 'one-stop-shop' of the best public 3D imagery data in the United States."

Get3Di can be accessed at get3di.com.

The initial release of Get3Di focuses on the state of Florida, other state datasets are coming online daily. If you are a public data provider and want more information on how your state can participate in Get3Di, please contact info@geocue.com.

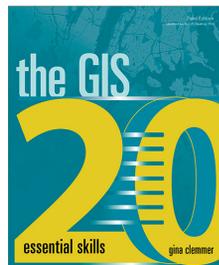
CALENDAR

- 11-15 August, **SPIE—Imaging Spectrometry XXIII**, San Diego, California. For more information, visit spie.org/OP423.
- 17-18 September, **GIS IN THE ROCKIES**, Denver, Colorado. For more information, visit <http://gisintherockies.org>.
- 28 September – 2 October, **GIS-PRO 2019**, New Orleans, Louisiana. For more information, visit www.urisa.org/gispro2019.
- 6-11 October, **Pecora 21/ISRSE 387**, Baltimore, Maryland. For more information, visit <http://www.asprs.org/event/pecora21-isrse387>.

FEATURES



405 An Interview— Dr. Qassim A. Abdullah



411 Book Review—*The GIS 20: Essential Skills, 3rd Edition*

COLUMNS

407 Mapping Matters

409 Grids and Datums

This month we look at the Swiss Confederation.

ANNOUNCEMENTS

412 Headquarters News

412 New ASPRS Members

Join us in welcoming our newest members to ASPRS.

DEPARTMENTS

402 Industry News

402 Calendar

410 Ad Index

434 ASPRS Sustaining Members

444 Who's Who in ASPRS

463 ASPRS Media Kit

PEER-REVIEWED ARTICLES

413 Special Issue Introduction

Fabio Remondino and Isabella Toschi

415 CNN-Based Dense Image Matching for Aerial Remote Sensing Images

Shunping Ji, Jin Liu, and Meng Lu

The performance of three typical CNN models is evaluated, when applied to deep learning based dense image matching of aerial stereo images.

425 Semantic Façade Segmentation from Airborne Oblique Images

Yaping Lin, Francesco Nex, and Michael Ying Yang

A Fully Convolutional Network (FCN) applied to oblique airborne images in order to perform semantic façade segmentation is presented.

435 RoofN3D: A Database for 3D Building Reconstruction with Deep Learning

Andreas Wichmann, Amgad Agoub, Valentina Schmidt, and Martin Kada

A novel 3D point cloud training dataset, named RoofN3D, is presented to train machine learning models for different tasks in the context of 3D building reconstruction.

445 Through-Water Dense Image Matching for Shallow Water Bathymetry

Gottfried Mandlburger

An investigation on Dense Image Matching for through-water DEM generation and the comparison with topo-bathymetric Lidar data.

455 Object-Based Point Cloud Analysis for Landslide and Erosion Monitoring

Andreas Mayr, Martin Rutzinger, and Clemens Geitner

A new method for semantically-supported and multi-temporal landslide monitoring based on TLS point clouds is presented.

See the Cover Description on Page 404

COVER DESCRIPTION



If the jagged structures sticking out of the flat, dry landscape in northwest New Mexico seem otherworldly, that's for a good reason. Protruding from the high desert plain, Shiprock has much in common with terrain on Mars, where volcanic activity formed similar razor-like, vertical walls.

The natural-color image above, acquired by the Operational Land Imager (OLI) on Landsat 8, shows Shiprock on April 12, 2017. Its central formation, a stone tower, consists of a 30 million-year-old volcanic neck, "the central feeder pipe of larger volcanic landform which has since eroded away," according to the New Mexico Bureau of Geology and Mineral Resources.

Shiprock rises 500 meters (more than 1,600 feet) above the surrounding desert, but there was a time when it hid below the soil like a sunken ship under water. The Bureau estimates that it formed 750 to 1000 meters (2,460 to 3,280 feet) below the Earth's surface. The formation was likely created in a hydrovolcanic eruption, in which "magma in upward-migrating dikes encounters groundwater, and heats it to steam under confining pressure," according to a paper on the Navajo Volcanic Field. Two perpendicular walls—"radial dikes" that radiate outward from the volcano's center—come together at the rocky outcrop. Initially, hot lava seeped into cracks in the older rock here, creating dikes. Since then, erosion has worn away the sandstone and shale around the dikes.

The red planet's ridges are thought to have been formed by similar volcanic and erosional processes to those that created Shiprock. "The features on Mars could be intrusive dikes like Shiprock," said Laszlo Kestay, director of the Astrogeology Science Center for the U.S. Geological Survey. "The region has plenty of volcanism and the Medusae Fossae Formation is easily eroded, making it a good host-rock for such features."

The origin of Mars's Medusae Fossae continues to puzzle geologists. Located in the Amazonis Planitia region of the planet, it lies between two volcanic domes, Tharsis and Elysium. Scientists generally agree that the formation is young and fine-grained, but beyond that, they are divided in theories on how it was created. Some hypotheses propose that it was formed as result of deposits from a meteorite impact into an aquifer underneath the planet's surface. Others suggest that it was created by pyroclastic flows—the outpouring of hot ash, lava, and other materials that run downhill after an eruption.

The inset image on the cover shows polygon-shaped ridges in the Gordii Dorsum portion of the Medusae Fossae region of Mars. The image was captured on April 9, 2010, by the High Resolution Imaging Science Experiment (HiRISE) camera on NASA's Mars Reconnaissance Orbiter. The dikes here are a fraction of a kilometer long—less than ten times shorter than their counterparts at Shiprock.

"For Shiprock, it is the classic location to show what is going on under the vents for a volcano," Kestay said. Although he is not aware of any studies directly on Shiprock that have been applied to the Red Planet, "It is the best example of a process that undoubtedly happens on Mars."

NASA Earth Observatory images by Jesse Allen, using Landsat data from the U.S. Geological Survey and Mars Reconnaissance Orbiter High Resolution Imaging Science Experiment (HiRISE) image courtesy of the University of Arizona Lunar and Planetary Laboratory. Story by Pola Lem.

For more information visit, <https://landsat.visibleearth.nasa.gov/view.php?id=90221>.

PHOTOGRAMMETRIC ENGINEERING & REMOTE SENSING



JOURNAL STAFF

Publisher ASPRS

Editor-In-Chief Alper Yilmaz

Assistant Editor Jie Shan

Assistant Director — Publications Rae Kelley

Electronic Publications Manager/Graphic Artist Matthew Austin

Photogrammetric Engineering & Remote Sensing is the official journal of the American Society for Photogrammetry and Remote Sensing. It is devoted to the exchange of ideas and information about the applications of photogrammetry, remote sensing, and geographic information systems. The technical activities of the Society are conducted through the following Technical Divisions: Geographic Information Systems, Photogrammetric Applications, Lidar, Primary Data Acquisition, Professional Practice, and Remote Sensing Applications. Additional information on the functioning of the Technical Divisions and the Society can be found in the Yearbook issue of *PE&RS*.

Correspondence relating to all business and editorial matters pertaining to this and other Society publications should be directed to the American Society for Photogrammetry and Remote Sensing, 425 Barlow Place, Suite 210, Bethesda, Maryland 20814-2144, including inquiries, memberships, subscriptions, changes in address, manuscripts for publication, advertising, back issues, and publications. The telephone number of the Society Headquarters is 301-493-0290; the fax number is 225-408-4422; web address is www.asprs.org.

PE&RS. *PE&RS* (ISSN0099-1112) is published monthly by the American Society for Photogrammetry and Remote Sensing, 425 Barlow Place, Suite 210, Bethesda, Maryland 20814-2144. Periodicals postage paid at Bethesda, Maryland and at additional mailing offices.

SUBSCRIPTION. For the 2019 subscription year, ASPRS is offering two options to our *PE&RS* subscribers — an e-Subscription and the print edition. E-subscribers can plus-up their subscriptions with printed copies for a small additional charge. Print and Electronic subscriptions are on a calendar-year basis that runs from January through December. We recommend that customers who choose both e-Subscription and print (e-Subscription + Print) renew on a calendar-year basis. The new electronic subscription includes access to ten years of digital back issues of *PE&RS* for online subscribers through the same portal at no additional charge. Please see the Frequently Asked Questions about our journal subscriptions.

The rate of the e-Subscription (digital) Site License Only for USA and Foreign: 1000.00 USD; e-Subscription (digital) Site License Only for Canada*: 1049.00 USD; Special Offers: e-Subscription (digital) Plus Print for the USA: 1365.00 USD; e-Subscription (digital) Plus Print Canada*: 1424.00 USD; e-Subscription (digital) Plus Print Outside of the USA: 1395.00 USD; Printed-Subscription Only for USA: 1065.00 USD; Printed-Subscription Only for Canada*: 1124.00 USD; Printed-Subscription Only for Other Foreign: 1195.00 USD. *Note: e-Subscription/Printed-Subscription Only/e-Subscription Plus Print for Canada include 5% of the total amount for Canada's Goods and Services Tax (GST #135123065). **PLEASE NOTE: All Subscription Agencies receive a 20.00 USD discount.**

POSTMASTER. Send address changes to *PE&RS*, ASPRS Headquarters, 425 Barlow Place, Suite 210, Bethesda, Maryland 20814-2144. CDN CPM #40020812

MEMBERSHIP. Membership is open to any person actively engaged in the practice of photogrammetry, photointerpretation, remote sensing and geographic information systems; or who by means of education or profession is interested in the application or development of these arts and sciences. Membership is for one year, with renewal based on the anniversary date of the month joined. Membership Dues include a 12-month electronic subscription to *PE&RS*. Or you can receive the print copy of *PE&RS* Journal which is available to all member types for an additional fee of 60.00 USD for shipping USA, \$65 USD for Canada, or 75.00 USD for international shipping. Dues for ASPRS Members outside of the U.S. will now be the same as for members residing in the U.S. Annual dues for Regular members (Active Member) is 150 USD; for Student members 50 USD for USA and Canada 58 USD; 60 USD for Other Foreign members. A tax of 5% for Canada's Goods and Service Tax (GST #135123065) is applied to all members residing in Canada.

COPYRIGHT 2019. Copyright by the American Society for Photogrammetry and Remote Sensing. Reproduction of this issue or any part thereof (except short quotations for use in preparing technical and scientific papers) may be made only after obtaining the specific approval of the Managing Editor. The Society is not responsible for any statements made or opinions expressed in technical papers, advertisements, or other portions of this publication. Printed in the United States of America.

PERMISSION TO PHOTOCOPY. The appearance of the code at the bottom of the first page of an article in this journal indicates the copyright owner's consent that copies of the article may be made for personal or internal use or for the personal or internal use of specific clients. This consent is given on the condition, however, that the copier pay the stated per copy fee of 3.00 USD through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, Massachusetts 01923, for copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to other kinds of copying, such as copying for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale.

DR. QASSIM ABDULLAH



Dr. Qassim Abdullah, Ph.D., CP, PLS, is an accomplished scientist with more than 40 years of combined industrial, research and development, and academic experience in analytical photogrammetry,

digital remote sensing, and civil and surveying engineering. His current responsibilities include designing and managing strategic programs to develop and implement new remote sensing technologies focused on meeting the evolving needs of geospatial users.

Currently, Abdullah is the Chief Scientist for Woolpert Geospatial Services and a member of Woolpert Labs team. In addition, he serves as an adjunct professor at the University of Maryland, Baltimore County, and at Penn State University, teaching graduate courses on unmanned aircraft systems (UAS), photogrammetry and remote sensing.

His latest accomplishments include evaluating and introducing Geiger and single photon lidar to the geospatial industry and leading Woolpert research activities in the field of UAS, its sensor calibration, and its workflow development.

He was elected as an ASPRS Fellow in 2017 and he is the recipient of several prestigious ASPRS awards, such as the Lifetime Achievement Award, the Photogrammetric (Fairchild) Award, the Outstanding Service Award for publishing the monthly column "Mapping Matters" for more than 10 years, the Presidential Citation Award in recognition to his contributions in co-authoring the new "Positional Accuracy Standards for Digital Geospatial Data," and the ASPRS Outstanding Workshop Instructor award.

He is an ASPRS-certified photogrammetrist and a licensed professional surveyor and mapper with the states of Florida, Oregon, Virginia, and South Carolina. He is also a certified thermographer by the FLIR Infrared Training Center and a Certified GEOINT Professional in Remote Sensing and Imagery Analysis (CGP-R) by the United States Geospatial Intelligence Foundation (USGIF).

What has been your most fulfilling accomplishment as a scientist / engineer? Why?

My most fulfilling accomplishment so far has been authoring the new "ASPRS Positional Accuracy Standards for Digital Geospatial Data." The last standards were published in the early 1980s, so this was truly a lifetime opportunity. I was glad that I was ready and well qualified for the task. It provided a constructive and impactful conduit for the theoretical and practical knowledge I have learned over 40 years of professional experience.

What significant changes have you seen in your field during your career?

The digital revolution has been the most significant change that I've experienced, and I have embraced it. It has introduced an exciting era that has transformed our business. It transitioned us from an industry relying solely on film cameras and analogue stereo plotters to the new era of digital cameras, lidar, GPS, IMU, UAS, IFSAR, multispectral and hyperspectral sensors, machine learning, image processing software, big data, cloud hosting and processing, etc. Truly, there are no limits to our new capabilities when it comes to sensor design and data processing.

What would you consider to be the most important advice you could offer a younger, upcoming scientist/engineer?

My sincere advice for young professionals is to embrace changes in technology and be prepared for difficult challenges through hard work, the pursuit of knowledge and a strong work ethic. Knowledge is not a status you reach and conquer; it is an evolving process that does not age or retire. Practice your profession with integrity and selflessness and be there for others when they need you.

What do you think are the most pressing scientific needs that should be met in the coming years? What would you like to see scientific research accomplish?

Data analytics and data mining are going to be our biggest challenges. As big data gets even bigger with the advancement of smart sensors, smart transportation, smart cities, smart infrastructure and smart everything, the need for analytical tools and algorithms grows even stronger. Mining big data is the way to improve citizens' lives and the environment we live in.

Have you seen many changes in ethical conduct within your field during your career?

The change and ease of use of geospatial sensors and processing software have drawn a new generation of mapping professionals, some of whom do not fully understand the profession or the ethical commitments that come with it. For example, the community of UAS-operators-turned-mappers. Due to UAS affordability and the ease of use of the processing software, some individuals started practicing mapping activities. Often, they try to sell products with false accuracy statements derived from their ignorance and detour from the reality of mapping practices. I advise them to invest in themselves and gain the necessary knowledge they need to qualify them for the title of mapping professionals.

AN INTERVIEW

How important do you think interdisciplinary collaboration will be for solving some of the challenges that lie ahead in science?

We are witnessing a time of great scientific achievement and vastly improved channels for communicating thoughts and capabilities. The internet of things (IoT) made it possible for any person, with the click of a mouse, to search a bank of giant libraries on the other side of the world. Such open and easy communication channels have made it possible for scientists and researchers from different disciplines to exchange ideas and thoughts and therefore cooperate toward greater innovation. A good example on the interdisciplinary collaboration is the introduction of UAS into our industry. This presented a challenge to our photogrammetric community and the current tools we are using. Because of the cooperation between computer scientists, machine learning techniques and photogrammetrists, a new generation of algorithms and processing software were developed to solve the challenges we are faced with in processing UAS-based imagery. It is always beneficial to investigate any problem from different perspectives to enrich the experience and to develop a creative solution.

What is the biggest open question in your field of photogrammetry that will require the most attention in the future?

I would not call it a question, but one of the biggest challenges we are against or will be faced with is the need for machine learning-based methods in data mining and data analytics. Today's space-based, ground and aerial sensors are acquiring massive amounts of data, but we can only utilize a portion of that data through our manual analyses and interpretation routines. The intelligence community is already in dire need of new ways of information extraction. The industry must pay careful attention to this need by allocating enough resources and grants for universities and research institutes to develop smart and automated methods of information extraction.

Tell us about your educational background including your doctoral research.

I obtained my bachelor's degree in civil engineering from the University of Basrah, Iraq. I came to the United States to complete my master's and doctoral degrees in photogrammetry and geodetic engineering at the Civil Engineering Department of the University of Washington, Seattle.

Do you have a particular teacher or professor who inspired your love of science? Why?

I've had several of them over the years. In my youth, my father was my mentor and biggest influence when it came to my interest in how things work. He was a self-made inventor and he got me interested in civil engineering when I helped him with his construction contracting business. I am also grateful to my PhD program supervisor, the late Dr. Sandor Veress of the University of Washington. In more recent years, I have had great admiration for my friend and colleague, Dr. Riadh Munjy

of CSU-Fresno, whom I consider a role model and a mentor when it comes to learning. He is a genius in applied mathematics and geodetic sciences, including photogrammetry

What is the focus of your current research?

My focus is on anything that involves sensors and product quality and accuracy. After GPS-controlled aerial triangulation, lidar continues to fascinate me. Today, I spend a lot of my research time on enhancing the quality of UAS-derived products and thermal survey for energy modeling and analysis.

Tell us about something we might see in our daily lives that directly correlates to your work.

These days, a lot of people, articles, and specifications reference the new "ASPRS Positional Accuracy Standards for Digital Geospatial Data." If you see that, it should remind you of me. I am proud to be part of the team that authored it.

Give us an example of how multi-disciplinary research directly contributed to your work.

My research on energy modeling and analysis using aerial thermal cameras combined our mapping methods and techniques with that of mechanical engineering through the joint research project I pursued with students and faculty of the University of Dayton. Woolpert developed the "Heat Score Map" to educate consumers of power companies on the heat efficiency of their homes.

What has your ASPRS membership meant to you?

ASPRS membership has provided me with a natural environment for my professional and technical development. Through the ASPRS journal, publications, webinars and technical conferences, I have found opportunities to polish my skills and knowledge.

Has ASPRS helped further your career? If so, how?

ASPRS offered me the opportunity to publish my monthly column "Mapping Matters." Answering the questions, I receive through the column has pushed me to read, investigate and learn how to provide solutions or ideas. When you are out there entertaining readers' questions, you need to be ready to provide guidance and advice to the people who seek it. Such guidance and advice must be communicated precisely, accurately and in a timely manner. There is a lot at stake when you claim to be an authority on any subject; you must earn respect and repeatedly prove yourself. The column and that presence helped my career because it pushed me to excel.

When you're not working on your research, what do you do in your free time?

I enjoy running in my free time. It's a great way to meet new people, and I enjoy the challenge of trying to improve my times or increase my distance. I'm fortunate to have several great places to run near my home. I also enjoy hiking, gardening, cooking, and artisan bread baking when I have the time.



MAPPING MATTERS

YOUR QUESTIONS ANSWERED

The layman's perspective on technical theory and practical applications of mapping and GIS

BY Qassim A. Abdullah, Ph.D., PLS, CP**

QUESTION:

Question: I am a college student working on my bachelor degree in Spatial Science (Surveying).

I am interested in photogrammetry and my study is on drone-based surveys. I have the following questions on the new "ASPRS Positional Accuracy Standards for Digital Geospatial Data:"

1. In sections 7.7 and 7.8, when it talks about checkpoint and ground control accuracy being $1/2 RMSE_{map}$, is $RMSE_{map}$ the desired/intended accuracy class?
2. With GCPs having three times the accuracy of the geospatial data set being tested, does that mean the GCP accuracy will be three times more accurate than the desired/intended accuracy class?
3. Do you use Table D.1 to calculate all the statistics and then use the results to determine the ASPRS accuracy class? Is that the typical workflow? Is there a sample report you can supply?
4. Are there guidelines on what you should aim for regarding the additional statistics discussed on the standards, such as max, min, skew, kurtosis and mean absolute?
5. Can you clarify what it means when you can state 'tested to meet' versus 'produced to meet'?
6. Can you direct me to a document regarding planning and best practice guidelines?
7. Do the vegetated area ground control points simply go on the bare ground between vegetation?
8. How do you assess seamline mismatch?
9. If an orthophoto fails a column in Table B.3 (e.g. the $RMSE_r$ is okay, but the accuracy at 95% CI exceeds the limit), do you select the accuracy class where your project meets or exceeds all standards in a single row?
10. I note that many drone-based surveys seem to have a mean error much higher than 25% of the RMSE. What does this information tell you about the quality of the project, and how can you correct it?
11. I also noticed that nearly every drone software company reports accuracy as a function of GSD, e.g. heights within three times the GSD. How are these related, and is GSD really related to accuracy in any way? I did a project with a GSD of 1 cm, but I achieved 11 mm RMSE heights and mean of 3 mm.
12. Can you direct me to where I can read more about rigorous total propagated uncertainty regarding photogrammetry?

James Wallace
University of Southern Queensland, Australia

Dr. Abdullah: Due to the length of this list of good questions, I will address them over the span of several articles.

PART I

Question 1—*In section 7.8 of the standards, when it talks about ground control accuracy being $1/2 RMSE_{map}$, is $RMSE_{map}$ the desired/intended accuracy class?*

Answer: Yes, it is. The standards, in section 7.8 impose the following relationship between the accuracy of the ground control points and the derived product or map:

- Accuracy of ground control designed for planimetric data (orthoimagery and/or digital planimetric map) production **only**:
 - $RMSE_x$ or $RMSE_y = 1/4 * RMSE_{x(Map)}$ or $RMSE_{y(Map)}$,
 - $RMSE_z = 1/2 * RMSE_{x(Map)}$ or $RMSE_{y(Map)}$
- Accuracy of ground control designed for elevation data, or planimetric data and elevation data production:
 - $RMSE_x$, $RMSE_y$ or $RMSE_z = 1/4 * RMSE_{x(Map)}$, $RMSE_{y(Map)}$ or $RMSE_z(DEM)$

Therefore, if your photogrammetric processing is solely for producing 2D planimetric data,

"The independent source of higher accuracy for checkpoints shall be at least three times more accurate than the required accuracy of the geospatial data set being tested."

Photogrammetric Engineering & Remote Sensing
Vol. 85, No. 6, June 2019, pp. 407–408.
0099-1112/19/407–408

© 2019 American Society for Photogrammetry and Remote Sensing
doi: 10.14358/PERS.85.6.407

such as orthoimagery and/or 2D vector maps with a horizontal accuracy class of 30 cm, your ground control points need to be surveyed with:

- Horizontal accuracy of $RMSE_x$ or $RMSE_y = 7.5$ cm
- Vertical accuracy of $RMSE_z = 15$ cm

However, if you plan to produce any elevation data, such as point clouds, contours or 3D vector maps, your ground control points need to be surveyed with:

- Horizontal accuracy of $RMSE_x$ or $RMSE_y = 7.5$ cm
- Vertical accuracy of $RMSE_z = 7.5$ cm

Question 2—In section 7.9 of the standards, when it talks about checkpoints being three times the accuracy of the geospatial data set being tested, does that mean the GCP accuracy will be three times more accurate than the desired/intended accuracy class?

Answer: Yes, it does. The standards in section 7.9 state the following:

“The independent source of higher accuracy for checkpoints shall be at least three times more accurate than the required accuracy of the geospatial data set being tested.” Therefore, if you are testing products that must meet a vertical accuracy class of 10 cm, your checkpoints should be surveyed to a vertical accuracy of $RMSE_z = 3.33$ cm.

Question 3—Do you use Table D.1 to calculate all the statistics and then use the results to determine the ASPRS accuracy class? Is that the typical workflow? Is there a sample report you can supply to me?

Answer: Table D.1 in the standards represents the methodology reported in the “National Standard for Spatial Data Accuracy (NSSDA)” testing guidelines. While you can use any organization to compute the statistical terms needed to calculate the accuracy figures (according to ASPRS standards), Table D.1 is particularly helpful.

“‘PRODUCED TO MEET’ statement is usually provided by data producers or providers when accuracy is not verified by an independent set of checkpoints.”

Question 4—Are there guidelines on what you should aim for regarding the additional statistics discussed in the standards, such as max, min, skew, kurtosis and mean absolute?

Answer: No, not at this moment. However, textbooks and manuals on statistics can be consulted to derive desirable values based on the limits of RMSE and bias set by the standards. I will add your question to the list of issues catalogued for future enhancements to the standards.

Question 5—Can you clarify what it means when you can state, ‘tested to meet’ versus ‘produced to meet’?

“‘TESTED TO MEET’ statement is usually provided by data users or their consultants when accuracy is verified using an independent set of checkpoints.”

Answer: Formal testing statements were provided to users in section 7.12 of the standards. There are two types of statements for reporting product accuracy:

- 1) **‘PRODUCED TO MEET’:** This statement is usually provided by data producers or providers when accuracy is not verified by an independent set of checkpoints. The producers report their achieved accuracy based on confidence in their workflows and by the data fit to the control they used to calibrate the products.
- 2) **‘TESTED TO MEET’:** This statement is usually provided by data users or their consultants when accuracy is verified using an independent set of checkpoints. For product accuracy to be independently validated according to ASPRS standards, the test must satisfy the following conditions:
 - a. Independent checkpoints are ground control points that are not used in the calibration process during product generation. Check points can also be derived from existing datasets with known accuracy.
 - b. Checkpoints should be more accurate than the tested products. According to the standards, checkpoints should be at least three times more accurate than the tested product.
 - c. To make it a valid statistical sample, regardless of the project size, there should be at least 20 well-distributed checkpoints used in the accuracy assessment.

Question 6—Can you direct me to a document regarding planning and best practice guidelines??

Answer: Unfortunately, no such document exists at ASPRS for UAS-related activities. I co-instruct several workshops during the ASPRS annual conferences that you may find useful. In those workshops, we explore best practices based on past project experience and technical guidelines developed over the years from our standard mapping practices. However, I am working with the ASPRS UAS subcommittee to encourage them to develop such a document that we can include as an addendum in a future version of the standards.
(to be continued)

***Dr. Abdullah is S Chief Scientist and Senior Associate at Woolpert, Inc. He is ASPRS fellow and the 2010 recipient of the ASPRS Photogrammetric Fairchild Award.*

The contents of this column reflect the views of the author, who is responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the American Society for Photogrammetry and Remote Sensing and/or Woolpert, Inc.



& GRIDS & DATUMS

THE SWISS

CONFEDERATION

BY Clifford J. Mugnier, CP, CMS, FASPRS

The Grids & Datums column has completed an exploration of every country on the Earth. For those who did not get to enjoy this world tour the first time, *PE&RS* is reprinting prior articles from the column. This month's article on the Swiss Confederation was originally printed in 2001 but contains updates to their coordinate system since then.

The area now known as the Swiss Confederation was occupied by Helvetians who were conquered by the Romans; the southwest was invaded by Burgundians, and the northeast was invaded by Alamanni. In 1291 the Forest Cantons, or provinces of Uri, Schwyz, and Unterwalden, formed an anti-Hapsburg league that became the nucleus of the Confederation. The perpetual neutrality of Switzerland was guaranteed by international agreement in 1815 at the Congress of Vienna and again in 1919 by the Treaty of Versailles. Its present constitution was adopted in 1874. The highest peak of this Federal Republic is Monte Rosa at 4,638 m (15,217 ft). The Swiss Confederation shares borders with France, Germany, Italy, Austria, and the Principality of Liechtenstein.

G. H. Dufour, later to become a general, founded the Eidgenössisches Topographisches Bureau (Topographical Bureau) in Geneva in 1838. Dufour decided to use the "carte du jour" projection of Europe for the time which was the ubiquitous Ellipsoidal Bonne originally used for topographic mapping by Cassini himself during the Napoleonic Campaigns. The grid used for this Ellipsoidal Bonne has a Latitude of Origin (ϕ_0) = 46° 57' 06.02" N, a Central Meridian (λ_0) = 7° 26' 24.75" East of Greenwich (5° 06' 10.80" East of Paris), a False Easting of 600 km and a False Northing of 200 km. The Berne Observatory Datum was circa 1840, and the ellipsoid used was the Schmidt 1831 where $a = 6,376,804$ m, and $1/f = 302.02$. Although my notes show a false origin for this old grid, I suspect that the original use was with the traditional quadrant system. The false origin probably crept into use as the base was updated after the Rosenmund System was introduced in the 20th century. From 1845 to 1864, the publication of the first accurate map, known as the "Dufour Map," covering the whole of Switzerland was performed at the scale of 1:100,000;



the slopes were shown by hachures. By this time, the classical triangulation of Switzerland comprised 40 triangles that had been observed with Kern instruments, and the average error of a figure was 0.86 arc seconds. The Swiss triangulation calculations were based on the Bessel 1841 ellipsoid after 1863 where $a = 6,377,397.155$ m, and $1/f = 299.1528$. The office was transferred to Berne in 1868, and the publication of the original surveys at 1:25,000 (Swiss Central Plains) and at 1:50,000 (Alps) with contours was performed from 1870-1916. The Old Berne Observatory Datum of 1898 published an Astronomical Latitude (Φ_0) = 46° 57' 08.66" N, based on observations executed by E. Plantamour in 1875 and an Astronomical Longitude (Λ_0) = 7° 26' 22.5" East of Greenwich. The defining azimuth to station Rötiflüh was (α_0) = 11° 12' 05.24". The ellipsoid height and deflection of the vertical are not defined and therefore are forced to zero at the origin.

Photogrammetric Engineering & Remote Sensing
Vol. 85, No. 6, June 2019, pp. 409–410.
0099-1112/19/409–410

© 2019 American Society for Photogrammetry
and Remote Sensing
doi: 10.14358/PERS.85.6.410

In 1900, the national mapping agency was renamed the Eidgenössische Landestopographie. The vertical Datum was defined as Repère Pierre du Niton 1902, a large rock in the harbor of Geneva, where $H_0 = 373.600$ m (“Gebrauchshöhe”) with a connection to the tide gauge in Marseilles, France. M. Rosenmund, an engineer with the bureau, developed a new projection. The new system is an oblique conformal cylindrical double projection, similar in concept to what General Jean Laborde developed for Madagascar. For the grid of the Swiss National Maps, the value $\phi_0 = 46^\circ 57' 07.90''$ N was chosen based on more recent measurements (1937), and the Central Meridian (λ_0) = $7^\circ 26' 22.5''$ East of Greenwich. The radius of the Gaussian Sphere evaluated at the grid origin for the Bessel 1841 ellipsoid is $R = 6,378,815.9036$ m. The Grid Scale Factor at Origin (m_0) = 1.00072913843, and the false origin is the same as previously listed for the old Swiss Bonne Grid of 600 km and 200 km. Conformal doubles became the “carte du jour” projections of Europe during the early 20th century, and the cylinder, the cone and the plane were all used as developable surfaces. The “oblique” for this Swiss system is really a misnomer; it’s merely transverse at an oblique latitude. Laborde actually used the first oblique cylinder with a tilt at the origin, Rousilhe used the first oblique plane for the stereographic for most of his hydrographic surveys, and Krovak used the first oblique (tilted) cone for the Czech Republic. Rousilhe was the only system developed that became a widely used conformal double projection. The Rosenmund projection is truly unique in the world for a national grid. The combination of the Bern Observatory (horizontal) Datum of 1898 with the Pierre du Niton (vertical) Datum of 1902 or LN02, and the Rosenmund projection and grid of 1903 have collectively been known since as the “CH1903 System” (Convention Helvetica 1903 System), of Switzerland.

The use of aerial photographs for map making commenced in 1930, although terrestrial photogrammetry was already in common use. By 1938, the first 1:50,000 series of the entire country was completed, and the office moved to a new building in Wabern, near Berne in 1941. The Swiss National Map series was completed at the 1:25,000 scale in 1979, and the first use of satellite receivers for national surveys was in 1979. In 1980 the office was renamed the Bundesamt für Landes topographie, and the first Swiss National map sheet was digitally updated in 1989.

In 1988, a new network of 104 GPS station observations began, and the resultant adjustment has become the new national (terrestrial) reference system of Switzerland and is called the CHTRS95. The local reference frame realized for the old CH1903 Datum is called LV95. The “CH1903+” was held fixed at a new fundamental point, Zimmerwald Z_0 where $\phi_0 = 46^\circ 52' 42.27031''$ N, $\lambda_0 = 7^\circ 27' 58.41774''$ East, and $X_0 = 191,775.0616$ m, $Y_0 = 602030.7698$ m, all still referenced

to the Bessel 1841. The new point was chosen because the original location no longer exists, and the original coordinates of triangulation point Gurten were kept to maintain orientation. The deflection of the vertical is now defined at Zimmerwald Z_0 : $\zeta_0 = +2.64''$, $\eta_0 = +2.73''$, and $H_0 = 897.8408$ m. Transforming the CH1903+ Datum to the CHTRS95 Datum (WGS84 ellipsoid) then is accomplished by $\Delta X = +674.253$ m, $\Delta Y = +015.053$ m, $\Delta Z = +405.324$ m, according to the Swiss Federal Office of Topography.

The European Datum of 1950 was computed for Switzerland by the U.S. Army Map Service in the 1950s, and to transform from EU50 to WGS84, $\Delta X = -87$ m, $\Delta Y = -96$ m, $\Delta Z = -120$ m. To transform from EU79 to WGS84, $\Delta X = -6$ m, $\Delta Y = -98$ m, $\Delta Z = -119$ m. These parameters are according to NIMA’s TR 8350.2, 3 January 2000.

According to the Swiss Federal Office of Topography, the seven-parameter Datum shift from CH1903+ to WGS84 is: $\Delta X = +660.077$ m ± 4.055 m, $\Delta Y = +013.551$ m ± 4.816 m, $\Delta Z = +369.344$ m ± 3.914 m, $\alpha = 2.484$ cc ± 0.417 cc, $\beta = 1.783$ cc ± 0.455 cc, $\gamma = 2.939$ cc ± 0.411 cc, and $M = 1.000000566 \pm 0.000000052$. “Different applications of these transformation parameters, particularly in northern Switzerland, have shown that WGS84 coordinates can be computed for all of Switzerland from the national coordinates with an accuracy better than 1 meter (1 sigma).”

UPDATE

Significant geodetic activities have been performed in Switzerland since 2001, and all is reported in National Reports to the International Union of Geodesy and Geophysics (IUGG). The accuracy of the known geoid in Switzerland is considered known to better than a few centimeters, largely due to an enormous amount of observations for the deflection of the vertical (DOV). The European Reference Frame (EUREF) is supported and the country is in full cooperation with all of its neighbors.

- 1 http://www.bernese.unibe.ch/publist/publist_2015.php.
- 2 <https://iag.dgfi.tum.de/fileadmin/IAG-docs/NationalReports2007/Switzerland.pdf>.
- 3 <https://www.swisstopo.admin.ch/en/knowledge-facts/surveying-geodesy/reference-frames/transformations-position.html>.
- 4 File:///Users/cjmce/Downloads/ch1903wgs84_e.pdf.

The contents of this column reflect the views of the author, who is responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the American Society for Photogrammetry and Remote Sensing and/or the Louisiana State University Center for GeoInformatics (C⁴G).

This column was previously published in *PE&RS*.

Ad Index

Geomni, Inc.		Geomni.net/psm		Cover 4
NRO		www.nro.gov/About-the-NRO/Business-Opportunities		Cover 2

The GIS 20: Essential Skills is a well-structured practical book for new, entry-level, and experienced Geographic Information System (GIS) users covering basic skills including, finding and editing data, querying GIS maps, creating and sharing reports, and publishing maps.

The introduction part of the book has the instructions on downloading and installing ArcGIS desktop software and chapter data files. Twenty chapters in this book covers data collection, organization, and storage (shapefiles, projections, attribute tables, queries) data analysis (geocoding, geodatabases, reports) and data visualization (publishing and sharing maps).

The book covers key terms in the use of ArcGIS such as shapefiles, excel spreadsheets, and map projections. In Chapter 1, the author explains shapefiles, ArcMap interface, and essential tools to familiarize users with the program. Chapter 2 focuses on key elements of creating maps and layouts by working with layers, changing colors, creating labels, layouts and legends, and creating titles, using scale bars and north arrows. Chapter 3, Projecting Shapefiles, starts with a brief summary on understanding the basics of projections. In the application part, State Plane Coordinate System (SPCS) and Universal Transverse Mercator (UTM) are used. The author compares the two projections by creating a map of the USA. This example gives the user a better sense of different projections.

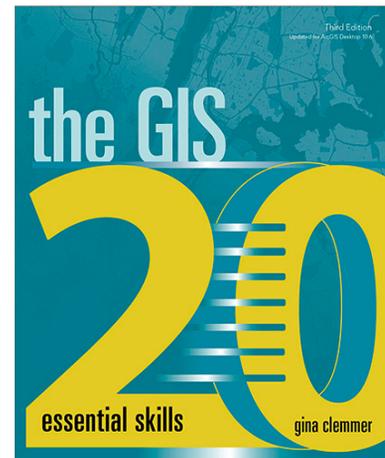
After covering the key terms, the author guides the reader by introducing basic concepts. Chapters 6 and 9, Creating Thematic and Creating a Categorical Maps focus on the visualization of written information.

Geocoding, another important skill is covered in detail in Chapter 8, gives the user the necessary skills to map anything with a physical location on a map. This chapter also covers address locators, auto and manual geocoding which gives the user the ability to convert non-spatial data into spatial. Chapter 10 explains the Global Positioning System (GPS) Point Mapping which is another method to generate addresses in GIS using latitude and longitude. Since our world is moving towards more practical and technological ways to collect data this chapter has significant importance.

Chapter 11, Editing, guides the user towards editing tasks such as changing boundary outline and creating shapefiles. This chapter also gives the user the chance to work with and understand vector data. Use of raster data is introduced later in Chapter 17, Working with Aerial Photography. Since the book uses both raster and vector data for various application it gives the user the chance to compare and see which data format is better for various applications.

Chapters 12, Creating Attribute Queries and 13, Creating Location Queries cover writing and erasing queries and creating new shapefiles based on query results. These chapters clearly explain the difference between attribute and location queries.

Chapter 14 covers GIS power tools for geoprocessing to manipulate spatial data such as buffer, merge, union, clip and dissolve. Chapter 15 explains explains ArcCatalog and creating



The GIS 20: Essential Skills, 3rd Edition

Gina Clemmer

Esri Press, Redlands, California. 2017. xiii and 182 pp., Illustrated, Softcover. ISBN 978-1-58948-512-9.

Reviewed by Esra Tekdal-Yilmaz, Assistant Professor, Surveying Engineering Program, The Pennsylvania State University, Lehman, Pennsylvania.

and working with geodatabases. It also covers the strengths of geodatabases and how they can make GIS data easier to manage and access.

Chapter 16 covers combining two shapefiles (and their data and tables) using spatial join function which is a very useful function that saves users time.

Chapter 18, Creating and Exporting Reports in ArcGIS and Chapter 19, Sharing Work using layer, map, and geoprocessing packages are guiding users to create and share reports and maps.

Chapter 20 covers publishing maps using ArcGIS online and creating geo-enabled PDFs to share GIS work with other GIS users. ArcGIS online is a popular cloud-based mapping and analysis solution hosted by ESRI which allows users to make maps, analyze data, and to share.

The twenty chapters mentioned above cover important materials and key points in GIS to guide the user through the learning process. Since instructions are supported by figures, tables, maps, and screenshots content is easy to follow. Overall, *The GIS20: Essential Skills* is a comprehensive, well-organized self-learning book presenting instructions on how to perform widely used 20 skills needed to successfully use ArcGIS Desktop software.

Photogrammetric Engineering & Remote Sensing
Vol. 85, No. 6, June 2019, pp. 411.
0099-1112/19/411

© 2019 American Society for Photogrammetry
and Remote Sensing

doi: 10.14358/PERS.85.6.411

JOURNAL STAFF

Editor-In-Chief

Alper Yilmaz, Ph.D., PERSeditor@asprs.org

Associate Editors

Rongjun Qin, Ph.D., qin.324@osu.edu
 Michael Yang, Ph.D., michael.yang@utwente.nl
 Petra Helmholtz, Ph.D., Petra.Helmholtz@curtin.edu.au
 Bo Wu, Ph.D., bo.wu@polyu.edu.hk
 Clement Mallet, Ph.D., clemallet@gmail.com
 Vasit Sagan, Ph.D., Vasit.Sagan@slu.edu
 Jose M. Pena, Ph.D., jmpena@ica.csic.es
 Xin Huang, Ph.D., huang_who@163.com
 Prasad Thenkabail, Ph.D., pthenkabail@usgs.gov
 Ruisheng Wang, Ph.D., ruishwang@ucalgary.ca
 Desheng Liu, Ph.D., liu.738@osu.edu
 Valérie Gouet-Brunet, Ph.D., valerie.gouet@ign.fr
 Yury Vizilter, Ph.D., viz@gosnias.ru
 Dorota Iwaszczuk, Ph.D., dorota.iwaszczuk@tum.de
 Qunming Wang, Ph.D., wqm11111@126.com
 Filiz Sunar, Ph.D., fsunar@itu.edu.tr
 Norbert Pfeifer, Ph.D., np@ipf.tuwien.ac.at
 Jan Dirk Wegner, Ph.D., jan.wegner@geod.baug.ethz.ch
 Hongyan Zhang, Ph.D., zhanghongyan@whu.edu.cn
 Zhenfeng Shao, Ph.D., shaozhenfeng@whu.edu.cn

Assistant Editor

Jie Shan, Ph.D., jshan@ecn.purdue.edu

Contributing Editors

Grids & Datums Column

Clifford J. Mugnier, C.P., C.M.S., cjmce@lsu.edu

Book Reviews

Sagar Deshpande, Ph.D., bookreview@asprs.org

Mapping Matters Column

Qassim Abdullah, Ph.D., Mapping_Matters@asprs.org

Sector Insight

Lucia Lovison-Golob, Ph.D., lucia.lovison@sat-drones.com

Project Management Column

Raquel Charrois, PMP, CP, CMS, PMP@asprs.org

ASPRS Staff

Assistant Director — Publications

Rae Kelley, rkelley@asprs.org

Electronic Publications Manager/Graphic Artist

Matthew Austin, maustin@asprs.org

Circulation Manager

Priscilla Weeks, pweeks@asprs.org

Advertising Sales Representative

Bill Spilman, bill@innovativemediasolutions.com

ASPRS MEMBERSHIP

ASPRS would like to welcome the following new members!

At Large

Maria M. Torres

Columbia Riviver

Nick Roberts

Florida

Chuckquan Dirod Harris

Mid-South

Dylan Buysse

Clark Jackson

Ryan Stephen Lynch

Mr. Christopher Morris

Austin Hogan Stone

Matthew Eric Thomas

New England

David A. Cook, III

Matthew Egan

Pacific Southwest

Steven Cyphers

Francis Staggs Hourigan

Potomac

Christopher Fox

Rocky Mountain

Dylan Addington

Western Great Lakes

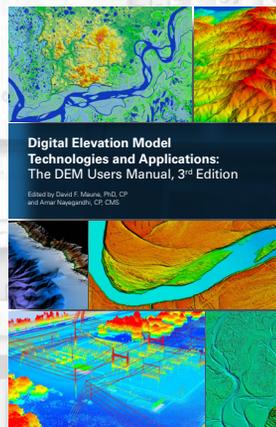
Michael Seidel, CMS

FOR MORE INFORMATION ON ASPRS MEMBERSHIP, VISIT

[HTTP://WWW.ASPRS.ORG/JOIN-NOW](http://www.asprs.org/join-now)

Your path to success in the geospatial community

NOW IN DIGITAL AND BOOK FORMAT!



DEM Users Manual, 3rd Ed.
 Hardback. 2018
 ISBN 1-57083-102-5
 Stock # 4959

Hardcopy

List Price \$100*
 ASPRS Member \$80*
 Student \$50*

Digital

eBook (Amazon.com) \$85

*Plus shipping/handling

This book is your guide to 3D elevation technologies, products and applications. It will guide you through the inception and implementation of the U.S. Geological Survey's (USGS) 3D Elevation Program (3DEP) to provide not just bare earth DEMs, but a full suite of 3D elevation products using Quality Levels (QLs) that are standardized and consistent across the U.S. and territories. The 3DEP is based on the National Enhanced Elevation Assessment (NEEA) which evaluated 602 different mission-critical requirements for and benefits from enhanced elevation data of various QLs for 34 Federal agencies, all 50 states (with local and Tribal input), and 13 non-governmental organizations. The NEEA documented the highest Return on Investment from QL2 lidar for the conterminous states, Hawaii and U.S. territories, and QL5 IfSAR for Alaska.

Co-Editors

David F. Maune, PhD, CP

Amar Nayegandhi, CP, CMS



Dewberry

[WWW.ASPRS.ORG](http://www.asprs.org)

Special Issue on the ISPRS TCII Symposium 2018

Fabio Remondino and Isabella Toschi, Guest Editors

This *PE&RS* Special Issue originates from the ISPRS Technical Commission II Midterm Symposium (www.isprs.org/tc2-symposium2018), that was held in Riva del Garda (Italy), in June 2018. ISPRS TCII “Photogrammetry” covers all aspects of measurement and information extraction in photogrammetry, with special focus on the geometric, radiometric and multi-temporal aspects of 3D imaging techniques. Particularly, the uniqueness of TCII mainly comes from its multi-scale perspective, encompassing theory and methods to derive accurate information from terrestrial, aerial and satellite images and point clouds, in various application fields – industrial metrology, heritage, geosciences, etc. These elements made the TCII Midterm Symposium a reference meeting point for researchers, practitioners and companies working in the field of photogrammetry, computer vision, geospatial data analysis and point cloud processing.

As event organizers, we selected the most promising contributions, that convey deep insight into the latest advances of photogrammetry, from image orientation and point cloud generation, to 3D scene reconstruction and large-scale machine learning methods for geospatial analysis. Data acquisition and processing in underwater and environmental applications are also treated. This Special Issue contains the most salient outcomes of the studies presented at the Symposium, and includes 10 peer-reviewed papers grouped into two issues.

In this issue, the reader will find the following articles:

CNN-Based Dense Image Matching for Aerial Remote Sensing Images

Dense stereo matching plays a key role in 3D reconstruction. The capability of using deep learning to stereo matching of remote sensing data is currently uncertain. This paper investigated the application of deep learning based stereo methods in aerial image series and proposed a deep learning based multi-view dense matching framework. Firstly, we applied three typical convolutional neural network models, MC-CNN, GC-Net and DispNet, to aerial stereo pairs, and compared the results with that of the SGM and a commercial software SURE. Secondly, on different datasets the generalization ability of each network is evaluated by using the direct transfer learning with models pretrained on other datasets and by fine-tuning with a small number of target training data. Thirdly, we present a deep learning based multi-view dense matching framework where the multi-view geometry

is introduced to further refine matching results. Three sets of aerial images as the main datasets, and two open-source sets of street images as auxiliary datasets are used for testing. Experiments show that: first, the performance of deep learning based stereo methods is slightly better than traditional methods. Second, both the GC-Net and MC-CNN have demonstrated good generalization ability and can obtain satisfactory results on aerial images using a pretrained model on several available stereo benchmarks. Third, multi-view geometry constraints can further improve the performance of deep learning based methods, which is better than that of the multi-view based SGM and SURE.

Semantic Façade Segmentation from Airborne Oblique Images

In this paper, oblique airborne images with very high resolution are used to address the problem from aerial views in urban areas. Traditional classification method (i.e. random forests) is compared with state-of-the-art fully convolutional networks (FCNs). Random forests use hand-craft image features including RGB, SIFT and Texton, and point cloud features consisting of normal vector and planarity extracted from different scales. In contrast, the inputs of FCNs are the RGB bands and the third components of normal vectors. In both cases, 3D features are projected back into the image space to support the facade interpretation. Fully connected conditional random field (CRF) is finally taken as a post-processing of the FCN to refine the segmentation results. Several tests have been performed and the achieved results show that the models embedding the 3D component outperform the solution using only images. FCNs significantly outperformed random forests, especially for the balcony delineation.

RoofN3D: A Database for 3D Building Reconstruction with Deep Learning

Machine learning methods, in particular those based on deep learning, have gained in importance through the latest development of artificial intelligence and computer hardware. However, the direct application of deep learning methods to improve the results of 3D building reconstruction is often not possible due, for example, to the lack of suitable training data. To address this issue, we present RoofN3D which provides a 3D point cloud training dataset that can be used to train machine learning models for different tasks in the context of 3D building reconstruction. The details about RoofN3D and the developed framework to automatically

derive such training data are described in this paper. Furthermore, we provide an overview of other available 3D point cloud training data and approaches from current literature in which solutions for the application of deep learning to 3D point cloud data are presented. Finally, we exemplarily demonstrate how the provided data can be used to classify building roofs with the PointNet framework.

Through-Water Dense Image Matching for Shallow Water Bathymetry

The introduction of Dense Image Matching (DIM) has reactivated the interest in photogrammetric surface mapping, as it allows the derivation of Digital Elevation Models (DEM) with a spatial resolution in the range of the ground sampling distance of the aerial images. While the primary field of application is wide-area mapping of topography and urban scenes, charting bathymetry of clear and shallow water areas is equally feasible via application of multimedia photogrammetry. The article specifically investigates the potential of through-water DIM for high resolution mapping of generally low textured shallow water areas using modern techniques like semi-global matching and off-the-shelf software. In a case study, the DIM-derived underwater surfaces of coastal and inland water bodies are compared to concurrently acquired laser bathymetry data. With an achieved penetration depth of more than 5 m and deviations in the dm-range compared to the laser data as reference, the results confirm the general feasibility of through-water DIM. However, sufficient bottom texture and favorable environmental conditions are a precondition for achieving accurate results.

Object-Based Point Cloud Analysis for Landslide and Erosion Monitoring

Today point clouds from close-range sensing are used for operational erosion and landslide monitoring. Distances between points from multi-temporal acquisitions can indicate surface deformation, while a designation of the underlying geomorphological processes is often handicapped by complex terrain structures and vegetation. We present an approach to landslide monitoring that integrates semantic information and 3D deformation detection automatically. Surface changes are assigned to (i) semantic object classes (landslide scarp, eroded area, deposit) and (ii) to spatially contiguous, individual objects (like parts of the landslide scarp and moving clods of turf and soil). We demonstrate this object-based approach with a time series of 13 topographic LiDAR point clouds, covering a site affected by shallow landsliding. The results of this case study illustrate how the presented methods translate the unstructured point clouds into information on geomorphological process dynamics to support erosion and landslide assessment.

CNN-Based Dense Image Matching for Aerial Remote Sensing Images

Shunping Ji, Jin Liu, and Meng Lu

Abstract

Dense stereo matching plays a key role in 3D reconstruction. The capability of using deep learning in the stereo matching of remote sensing data is currently uncertain. This article investigated the application of deep learning-based stereo methods in aerial image series and proposed a deep learning-based multi-view dense matching framework. First, we applied three typical convolutional neural network models, MC-CNN, GC-Net, and DispNet, to aerial stereo pairs and compared the results with those of the SGM and a commercial software, SURE. Second, on different data sets, the generalization ability of each network is evaluated by using direct transfer learning with models pretrained on other data sets and by fine-tuning with a small number of target training data. Third, we present a deep learning-based multi-view dense matching framework where the multi-view geometry is introduced to further refine matching results. Three sets of aerial images as the main data sets and two open-source sets of street images as auxiliary data sets are used for testing. Experiments show that, first, the performance of deep learning-based stereo methods is slightly better than traditional methods. Second, both the GC-Net and the MC-CNN have demonstrated good generalization ability and can obtain satisfactory results on aerial images using a pretrained model on several available stereo benchmarks. Third, multi-view geometry constraints can further improve the performance of deep learning-based methods, which is better than that of the multi-view-based SGM and SURE.

Introduction

Reconstructing terrestrial 3D scenes from stereo or multi-view aerial/satellite images has been a core problem in photogrammetry and remote sensing. The key technique is to obtain the correspondent points for each pixel in the stereo image pairs, which is commonly called dense stereo matching. A conventional process of dense stereo matching includes four steps (Scharstein and Szeliski 2002). The first is to calculate matching cost. Typical matching costs include luminance difference, correlation coefficient, and mutual information related to a pixel's values or distributions. Given a search area along an epipolar line, one of these costs is calculated pixel by pixel, and the minimum cost corresponds to the optimal matching candidate. These empirical designed costs could be heavily affected by the nontexture area, mirror reflection, and repeated pattern (Kendall *et al.* 2017). The second step is to aggregate the matching costs. The cost aggregation is commonly a weighted sum of all of the matching costs within a given neighborhood. However, traditional methods, such as Graph-Cut (Boykov *et al.* 2001) and SGM (Furukawa and Ponce 2010), could oversimplify the cost aggregation, for

example, using an empirical fixed window size and treating the pixels in a neighborhood independently. The third step is disparity calculation. At each pixel, the disparity value corresponds to the minimum cost. Interpolation can be used to achieve subpixel accuracy. The last step is parallax refinement including a series of postprocessing techniques, such as left and right consistency checking, median filter smoothing, and subpixel enhancement. Finally, the dense disparity map could be converted to a depth map to reconstruct a 3D scene.

Dense matching has been extensively studied. We classify them into conventional methods and deep learning-based methods. Graph Cut (Boykov *et al.* 2001) is a widely accepted global stereo matching algorithm introduced early in the 21st century. It uses graph theory, especially graph cut, to solve the problem of 2D energy minimization. Global matching algorithms such as Graph Cut are computationally expensive and are not suitable for large-volume remote sensing images. In 2008, a semiglobal matching method (SGM) with higher matching efficiency has been proposed (Hirschmuller 2008). The SGM considers the 2D cost aggregation as 16 1D cost aggregations and performs dynamic programming to solve the minimum cost. The Patch-Match (Bleyer *et al.* 2011) algorithm uses the local correlation of the image and assumes that the areas around the matching points also match each other. In addition to stereo, multi-view geometry is often used in dense matching for stronger constraints. The PMVS (Patch-Based Multi-View Stereo) algorithm (Furukawa and Ponce 2010) extracts feature points and retrieves the surrounding patches centered at the feature points and performs patch matching to obtain quasi-dense matching points. The SURE algorithm proposed by Rothermel *et al.* (2012) extends the stereo SGM to multi-view image matching. Multi-view geometry fusion is added to merge the redundant depth estimation values to achieve mutual restraint.

Deep learning, especially *convolutional neural networks* (CNNs), have been widely applied in image processing. It has been shown that CNN-based methods (LeCun *et al.* 2015; Schmidhuber 2015) can not only improve the accuracy of image recognition and classification but also increase the efficiency of online operations. More important, empirical and manual feature engineering can be replaced by the powerful representation learning ability of deep learning. From 2015, some studies have started applying deep learning to dense stereo matching to replace the empirically designed matching costs, cost aggregation, or the whole matching procedure. The matching results obtained on the computer vision benchmarks have gradually exceeded the traditional methods in speed and accuracy.

Two strategies are commonly used for CNN-based dense matching methods: (1) the end-to-end prediction from image to disparity image and (2) applying CNN to learn parts of the four standard steps of stereo matching. For example, the MC-CNN network (Zbontar and LeCun 2015) automatically learns

Shunping Ji and Jin Liu are with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China, 430079 (jishunping@whu.edu.cn).

Meng Lu is with the Department of Physical Geography, Faculty of Geoscience, Utrecht University, Utrecht, The Netherlands.

Photogrammetric Engineering & Remote Sensing
Vol. 85, No. 6, June 2019, pp. 415–424.
0099-1112/19/415–424

© 2019 American Society for Photogrammetry
and Remote Sensing
doi: 10.14358/PERS.85.6.415

the matching cost (i.e., step 1) via a Siamese CNN structure. SGM-Net (Seki and Pollefeys 2017) introduces a CNN learning penalty terms in the standard process of SGM.

The end-to-end learning strategy predicts disparity images directly from stereo pairs. For example, DispNet (Mayer *et al.* 2016) is a typical fully convolutional network (FCN). In the encoding stage, the network extracts high-level features of stereo images layer by layer; in the decoding stage, the network restores the feature map from coarse to the original image resolution to produce the disparity map. GC-Net (Kendall *et al.* 2017) makes full use of the geometric information and semantic information between pixels. 3D volume with context information, consisting of 2D feature maps cross disparities extracted by 2D CNN, is convoluted by a series of 3D kernels and finally flattened into a 2D disparity image. The pyramid stereo matching network (PSM-Net) (Chang and Chen 2018) is a pyramid stereo matching network consisting of spatial pyramid pools and 3D convolutional layers. It combines the global background information into stereo matching to achieve reliable estimation of occlusion areas, textureless areas, or pattern repeated areas. The cascade residual learning (Pang *et al.* 2017) method concatenates two improved DispNet networks. The first network obtains the initial disparity value between stereo pairs; the second network uses the residuals of the previous stage to train a finer disparity map. In Shaked and Wolf (2017), a new highway network structure is proposed; multi-level residual skip connections and composite loss function is applied. All of these methods operate in a supervised manner, requiring high-precision disparity maps as labels for training. An exception was presented by Zhong *et al.* (2017), who designed a CNN to learn disparity maps directly from the stereo pair without training samples based on the assumption that the left disparity map (based on the left image) and the right disparity map (based on the right image) learned by the network are inverted.

Although the deep learning-based stereo methods have been applied to match close-range images and have achieved improved results compared to conventional methods, it has not been applied to remote sensing images. In addition, the capability of it has not been compared with the mainstream photogrammetric algorithms. The main objective of this work is to comprehensively investigate the application of deep learning-based stereo methods on aerial remote sensing images and to compare them with conventional methods, including commercial software. The second contribution is to introduce multi-view geometry for deep learning-based dense matching for the first time. Multi-view geometry gives more constraints between co-visible images and could be more robust than a stereo vision. In addition, we evaluate the generalization ability of deep learning in aerial stereo matching; that is, can the model trained on the available benchmark data sets be directly applicable to aerial imagery? It is highly relevant, as in the existing and upcoming aerial image data sets, accurately labeled samples are commonly lacking or insufficient.

Method

CNN for Learning Only Matching Cost

The similarity score calculated by the normalized correlation coefficient, intensity difference, or cross entropy is shown to be incompetence in some complex situations. Matching cost learned by CNNs has shown advantages over those empirical designs on close-range data (Han *et al.* 2015; Zagoruyko and Komodakis 2015; Aguilera *et al.* 2016). In this study, we use the MC-CNN (especially the fast structure) (Zbontar and LeCun 2015) for evaluating the performance of a CNN-based matching cost.

The fast MC-CNN structure uses Siamese convolutional networks with shared weights to extract feature vectors from the input stereo tiles. The dot product operator measures the similarity between the two extracted and normalized feature vectors. The network structure is shown in Figure 1.

In this article, the number of convolution layers is set to 4, and the convolution kernel size is set to 3×3 . The MC-CNN calculates the loss values of a pair of positive and negative samples and trains the network by minimizing a hinge loss function. The hinge loss of the positive and negative samples is defined as $\max(0, m + s-, s+)$ where $s+$ is the output of the positive sample, $s-$ is the output of the negative sample, and the tolerance m is set to 0.2.

Based on the initial disparity map that is computed with the learned matching score, a series of postprocessing steps including cost aggregation (Zhang *et al.* 2009), semiglobal matching, left and right consistency check, subpixel enhancement, median filtering, and bilateral filtering are applied to ensure the best matching results. The process bears a resemblance to SGM.

Special CNN for End-to-End Stereo Matching

We use the GC-Net (Kendall 2017) as an end-to-end stereo benchmark algorithm and evaluate its performance on aerial data sets. Its core concept of geometry and context combination is to treat the parallax as the third dimension orthogonal

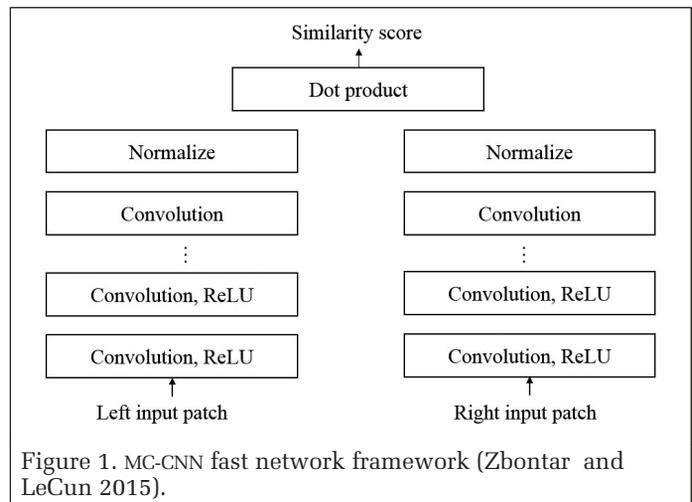


Figure 1. MC-CNN fast network framework (Zbontar and LeCun 2015).

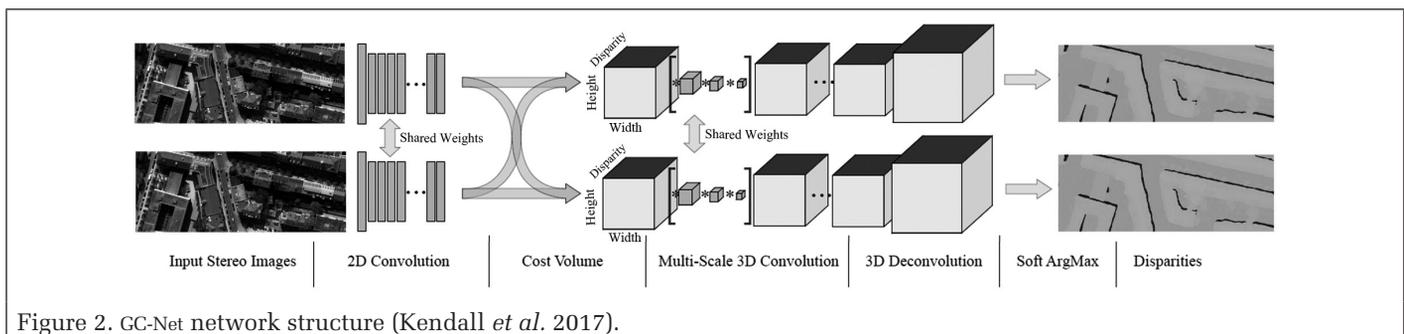


Figure 2. GC-Net network structure (Kendall *et al.* 2017).

to the image plane. The feature maps, which are learned from a series of 2D convolutional layers, consist of a 3D tensor across each disparity. The 3D convolution then learns geometric and semantic features to obtain an optimal disparity map (i.e., a curved surface cutting through the 3D tensors). The GC-Net network structure is shown in Figure 2. Several 2D convolutions are applied to the stereo pairs with shared weights to extract feature maps. The m feature maps of the last layer with size of $w \times h$ are concatenated across $0 \sim n$ disparity to constructed m feature volumes with size of $w \times h \times (n + 1)$. 3D convolution and deconvolution are executed to learn a series of 3D feature maps. The size of the final 3D maps is $W \times H \times n$, where H and W are the length and width of the original image, respectively. The last step is to flatten the 3D feature map to a 2D disparity map with a Soft Argmin operation. The predicted disparity map and its corresponding label are utilized to train the optimal parameters by iterating the forward- and backward-propagating process. In our experiment, we used 18 2D convolutional layers with $32 \ 3 \times 3$ kernels; 14 3D convolutional layers, and five 3D deconvolution layers with a kernel size of $3 \times 3 \times 3$.

General FCN for End-to-End Stereo Matching

Although GC-Net can combine context information with geometry and is preferable theoretically, the 3D convolution requires much more memory compared to the 2D convolution. An end-to-end learning from a 2D image pair to a 2D disparity map could be achieved using a general FCN. Among several 2D FCNs that have reported satisfactory performance in stereo matching (Mayer *et al.* 2016), we used the DispNet network for evaluation, which is based on the FlowNet (Flow Estimation Network) (Dosovitskiy *et al.* 2015) and is modified to calculate disparity maps. DispNet is a typical FCN structure consisting of an encoder and a decoder. The encoder has six convolutional layers where the kernel sizes of the first two layers are 7×7 and 5×5 , respectively, and all the other layers are 3×3 . The decoder consists of five up-convolutional layers

with kernel size 4×4 ; each layer is first concatenated with the feature map of the corresponding layer in the encoding step and then merged by a series of convolution operations. The structure is similar to mainstream semantic segmentation CNN, such as U-Net (Ronneberger *et al.* 2015). A schematic diagram of the DispNet network is shown in Figure 3.

Transfer Learning

The performance of applying a pre-trained dense matching model on a different data set is a key issue in practice, as we seldom have enough highly accurate disparity maps in target data to train a CNN with multiple parameters. We adopt and evaluate two transfer learning strategies (Pan and Yang 2010) direct model transfer and model fine-tuning—with small samples available in target data set.

A direct model transfer utilizes the model pretrained on the source data set, predicting the target data without any parameter tuning. This method requires high generalization ability of the model. Especially, we evaluate whether the learned features through a CNN for finding pixel correspondence is universal for all sorts of data sources, including close-range images, aerial images, and simulated scene images.

Assuming that the target data set have insufficient samples to train a robust network model, fine-tuning with these samples conditioned on a pretrained model is a common choice. It can reduce the number of iterations required for training a new model and alleviate the problem of insufficient samples. There are two different strategies for fine-tuning: one is to train the parameters of all layers, and the other is to train only the top layers and freeze the bottom layers. As the number of network layers involved in this article is relatively few, we retrain all the parameters in the CNN.

Deep Learning-Based Multi-View Dense Matching

Up to now, most of the CNN-based stereo methods evaluate their performances only at a pixel level (i.e., disparity rather than in depth), which is insufficient, as their accuracy is not

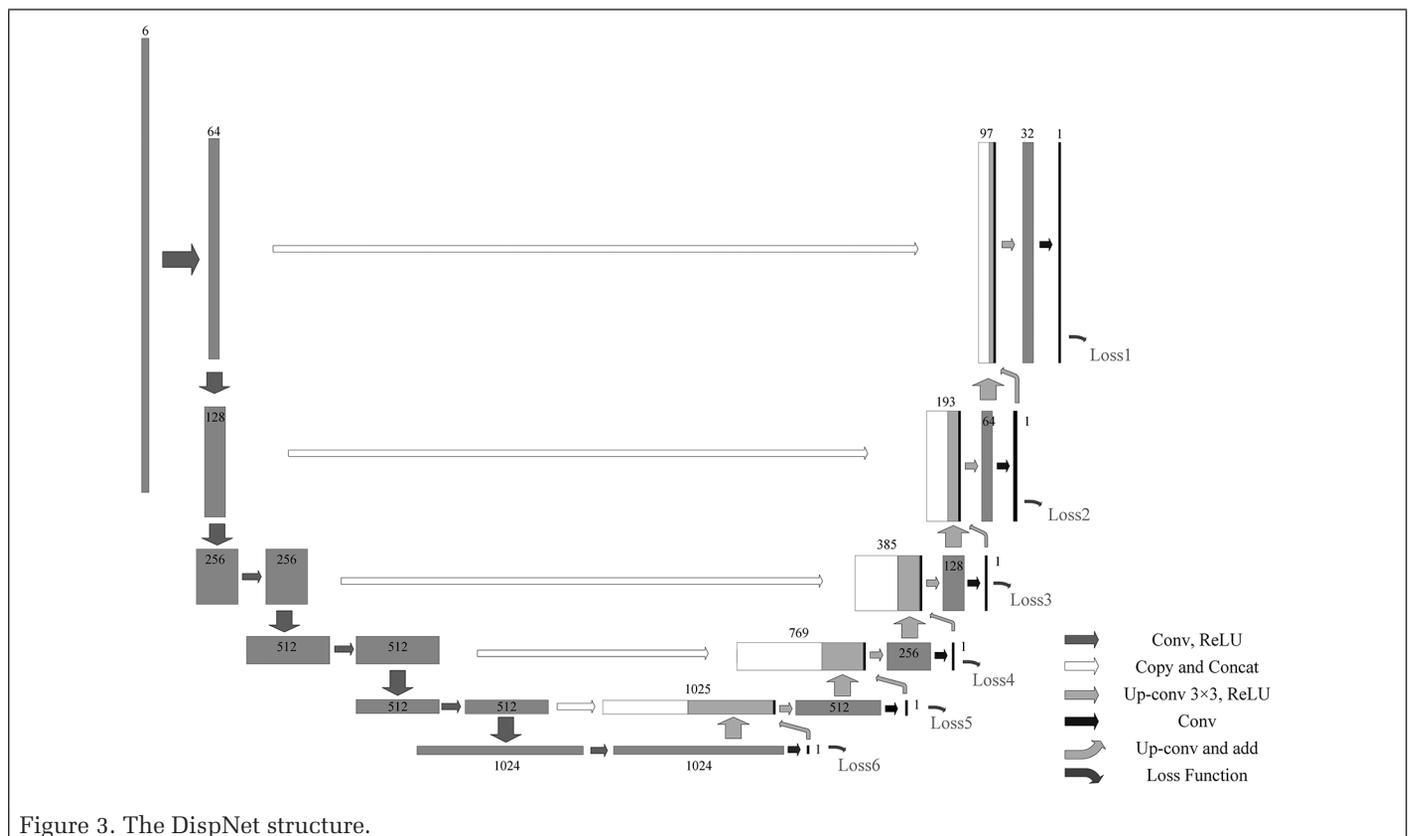


Figure 3. The DispNet structure.

equal. Further, the important multi-view geometry has not been applied to CNN-based stereo matching. We translate the disparity to depth and merge the multi-view matching results on the georeferenced coordinate system for comprehensively understanding the depth accuracy that a CNN-based method could reach.

To generate rectified epipolar stereos of each image pair given a base image, we adopt the method used in Fusiello *et al.* (2000), where only some simple 3×3 2D perspective transformation matrices are required for transferring between the epipolar stereos and between the original image and the rectified one. The relationship between the disparity and the depth of a pixel is

$$Z = \frac{Bf}{d} \quad (1)$$

where B denotes the baseline between the rectified stereos and d , f , and Z are the disparity, focal length, and depth value, respectively. Similarly, the depth of a pixel in a rectified image along the ray is

$$D_b^r = \frac{B\sqrt{(x_b^r)^2 + (y_b^r)^2 + f^2}}{d} \quad (2)$$

where (x^r, y^r, f) is the camera coordinate of a point in the rectified base image. The rectified depth map can be converted into the original image coordinate according to the inverse perspective transformation matrixes. In Figure 4, the depths of the ray in stereo (l, b) and stereo (r, b) are D_1 and D_2 , respectively. Ideally, D_1 and D_2 are equal. According to the average accuracy σ (e.g., 0.3 m) of the depth map obtained by stereo methods, a threshold could be set to discriminate whether the depth values of the two stereos are consistent.

Figure 5 shows the process of the deep learning-based multi-view dense matching. Triple-view images are used for demonstration where I_b represents the base image and I_l and I_r are the left image and the right image, respectively. The two stereos (consisting of I_l, I_b and I_r, I_b , respectively) are first rectified via the corresponding perspective transformation matrixes H for obtaining two epipolar stereos. Then a CNN model is separately applied to the two rectified stereos to output two corresponding disparity maps, $d_{1,b}^r$ and $d_{2,b}^r$. The two disparity maps are converted to depth maps $D_{1,b}^r$ and $D_{2,b}^r$ according to Equation 2. Then the depth maps are transferred to the original image coordinates (Figure 4). Finally, the difference between the values of every overlapped pixel below a given threshold indicating multi-view compatibility and the mean value is used as the final depth; otherwise, the pixel is treated

as an abnormal disparity point, which is filtered out and filled in with its neighborhood values.

In this work, we utilize the multi-view constraint as a postprocessing, as in the conventional methods (Rothermel *et al.* 2012), instead of learning a depth map directly from multi-view stereos. This multi-view constraint could also be embedded into the learning loop (forward and backward propagation) in the end-to-end deep learning methods. However, as is shown in Figure 5, many pixel-wise coordinate transformations and massive computations are involved in the process, which may not be feasible with a single graphics processing unit (GPU).

Data Sets

To evaluate the performance of the deep learning-based stereo methods in aerial images, five data sets (KITTI, Driving, Hangzhou, München, and Vaihingen) are used in this experiment, among which the KITTI and Driving data sets are open source and consist of close-range data. The Hangzhou data set is collected from unmanned aerial vehicles (UAVs) and the München and Vaihingen data sets from traditional aerial photography platforms.

KITTI Data Set

The KITTI data set (Menze and Geiger 2015) consists of street-scene data in the German city of Karlsruhe captured from stereo cameras with a 54-cm baseline mounted on the roof of a car. The ground-truth depth is recorded by a rotating a laser scanner installed behind the left camera, ensuring that 30% of

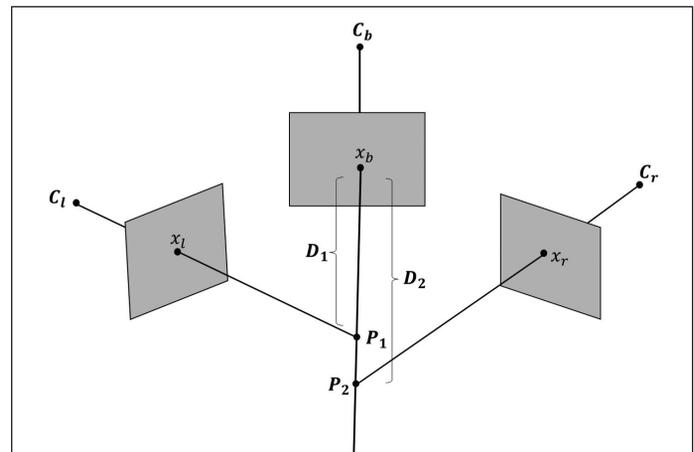


Figure 4. Depths of the ray in stereo (l, b) and stereo (r, b) are D_1 and D_2 in the uniform base image coordinate. The difference of D_1 and D_2 reflects the multi-view consistency.

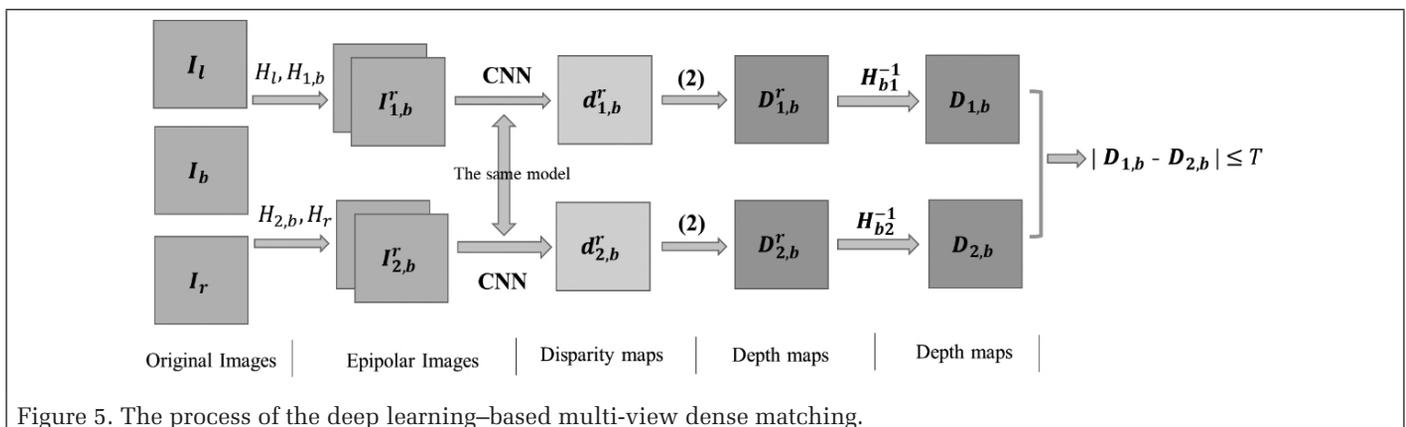


Figure 5. The process of the deep learning-based multi-view dense matching.

pixels have depth value. Both KITTI2012 and KITTI2015 data sets consist of about 200 rectified stereo pairs for training and 200 pairs with unpublic labels for testing with an average image size of 1240×375 pixels. In our experiment, eighty percent of the original training set was used for training and the remaining 20% for testing.

Driving Data Set

The Driving data set (Mayer *et al.* 2016) is a set of virtual street-view images with configured accurate stereo pairs, 3D scenes, and disparity maps. This data set consists of much more data than other existing real data sets and could facilitate the training of large CNNs. The rectified stereo image size is fixed at 960×540 . In the experiment, 300 rectified 960×540 stereo pairs were selected from the whole data set, from which 80% were used as training sets and the remaining 20% as test sets.

Hangzhou Data Set

The Hangzhou data set is a series of UAV images that were taken by a drone at a low altitude of about 640 m above-ground, recording the scene of a mountain village near the city of Hangzhou, China, in August 2017. Twenty aerial images consist of four strips with 80% heading overlap and 60% side overlap, with 0.07-m ground sampling distance (GSD). Lidar point cloud measurements are used as the ground truth with approximate 0.15-m height accuracy. The original 9000×6732 images are rectified into epipolar images. The disparity values of each pixel are calculated from the point cloud. Due to the limitation of the GPU capacity, the rectified aerial images are cropped into subimages of size 1325×354 . After manually removing some image pairs with large visual bias between LiDAR point and images, the remaining 328 pairs of images were used as training sets and 40 pairs as test sets.

München and Vaihingen Data Sets

The München aerial data set consists of 15 $14,114 \times 15,552$ aerial images that were captured in three strips with 80% heading overlap and 80% side overlap, covering urban buildings, roads, and greenbelts. The Vaihingen data set consists of 36 $9,420 \times 14,430$ aerial images of three strips with 60% heading overlap and 60% side overlap, covering flat rural scenes. The München and Vaihingen data sets have a GSD of 0.14 m and 0.23 m, respectively. The ground truth of depth is provided in forms of semidense DSM with 0.1 m and 0.2 m GSD, respectively, which was generated and filtered using the median of the results calculated by seven commercial software and shows high visual accuracy. It is empirically inferred that the DSM height accuracy is between 0.2 m and 0.4 m. Similar to the preprocessing of the Hangzhou data set, the rectified epipolar images are cropped into subimages of 1150×435 and 955×360 , respectively. Finally, the München data set consists of 540 stereo pairs, and the Vaihingen data set consists of 740 pairs. The ratio of the training and test sets is 4:1.

Experiment and Result Analysis

Three experiments were designed to comprehensively evaluate the performance and generalization ability of the deep learning method in aerial remote sensing images. The first tests the performance of deep learning methods using the three aerial data sets Hangzhou, München, and Vaihingen. The results are compared with the SGM and SURE. The second is to test the generalization of stereo dense matching. The models pretrained on computer vision open data sets were applied to aerial imagery. The third one is to extend deep learning-based stereo matching to multi-view matching and evaluate its performance.

Three-pixel-error (3PE, the percentage of pixels with a disparity error less than three pixels) and one-pixel-error (1PE, the percentage of pixels with a disparity error less than one pixel) are used for accuracy assessment. In multi-view matching, the depth map is compared to the ground truth in meters. All deep learning methods are implemented with an NVIDIA Titan Xp 12G GPU.

Comparison of the Deep Learning-Based Methods with Traditional Methods

We evaluated the performance of the MC-CNN, GC-Net, and DispNet on dense matching and compared this with the SGM and SURE. The basic settings of these methods/software are as follows.

For the MC-CNN, at the training stage, the input of model is 128 pairs of positive and negative samples composed of 9×9 image blocks. Small batch gradient descent is adopted to minimize the loss, and the momentum is set to 0.9. The learning rate is 0.002 and was adjusted to 0.0002 after several iterations.

For the GC-Net, which is less effective on the sparse disparity map, the network is trained only on three dense data sets (the data sets that were not processed are represented as “—” in Table 1). The batch size is set to 1, all data are iterated 50 times, and the learning rate is 0.001.

For the DispNet, the batch size is set to 32, and the learning rate is 0.0001 and was gradually decreased during the training process. All training data were iterated 1500 times.

For the SGM, we used the implemented function in Opencv3.0 library, with postprocessing, such as Gaussian smoothing and median filtering.

For the SURE, the inputs are the original aerial images and orientation information, and the output is a 3D model in the format of OSGB. Therefore, experiments were conducted on three sets of aerial images exclusively. We used the 3D workflow mode with the default parameters and settings. To evaluate the accuracy, the corresponding disparities of each point on epipolar images are calculated from the 3D model and compared with the real disparity values.

The performance of the traditional methods and deep learning methods on five data sets is shown in Table 1.

In general, the GC-Net performs the best; the MC-CNN performs comparable to the commercial software SURE in 3PE and slightly lower in the 1PE indicator and far superior to SGM; DispNet performs the worst in 1PE.

Using the GC-Net model, the accuracy on the flat Vaihingen data set is 99.7% (98.0%) and is slightly better than all the other well-performed methods. The München data set has obvious variations in elevation and has more discriminatory power for the comparison of the methods. The 3PE of the GC-Net is about 2% higher than the MC-CNN model (second best), and 1PE is 7.4% higher than the SURE (second best). On the Driving data set, 92.6% (85.7%) of the test accuracy is also much higher than other methods.

The performance of MC-CNN model on all the data sets is stable. The accuracy of each data set is much higher than the SGM: 6.7% (4.6%) higher in the KITTI2015 and 5.7% (7.7%)

Table 1. Accuracy comparison between the traditional and deep learning stereo methods on the five data sets.

Method	Accuracy (3PE/1PE)				
	KITTI2015	Driving	Hangzhou	München	Vaihingen
MC-CNN	0.960/0.778	—	0.953/0.816	0.965/0.867	0.992/0.932
GC-Net	—	0.926/0.857	—	0.984/0.953	0.997/0.980
DispNet	0.937/0.737	0.835/0.547	0.923/0.591	0.883/0.532	0.950/0.710
SGM	0.893/0.732	0.713/0.505	0.896/0.739	0.921/0.859	0.987/0.925
SURE	—	—	0.968/0.831	0.932/0.879	0.990/0.969

higher in the Hangzhou data sets. On the three aerial image data sets Hangzhou, München, and Vaihingen, the MC-CNN is equivalent to the multi-view-based SURE in 3PE and slightly lower (1.5%, 1.2%, and 3.7%, respectively) in 1PE.

The DispNet with a generic FCN structure obtained the worst accuracy on the remote sensing image data sets. The poor results on 1PE especially reflect the limitation of a generic model on dense matching tasks. DispNet is suitable only

for stereos with very small disparity values. For example, DispNet preforms better than SGM on the KITTI2015 data set, which has a maximum disparity of merely 70 pixels, while on remote sensing data sets with large terrain fluctuations, the results become inaccurate and unstable.

Figure 6 shows the predicted disparity maps of all of the methods on the three aerial image data sets. From top to bottom are the stereo image pairs, ground truth, and the

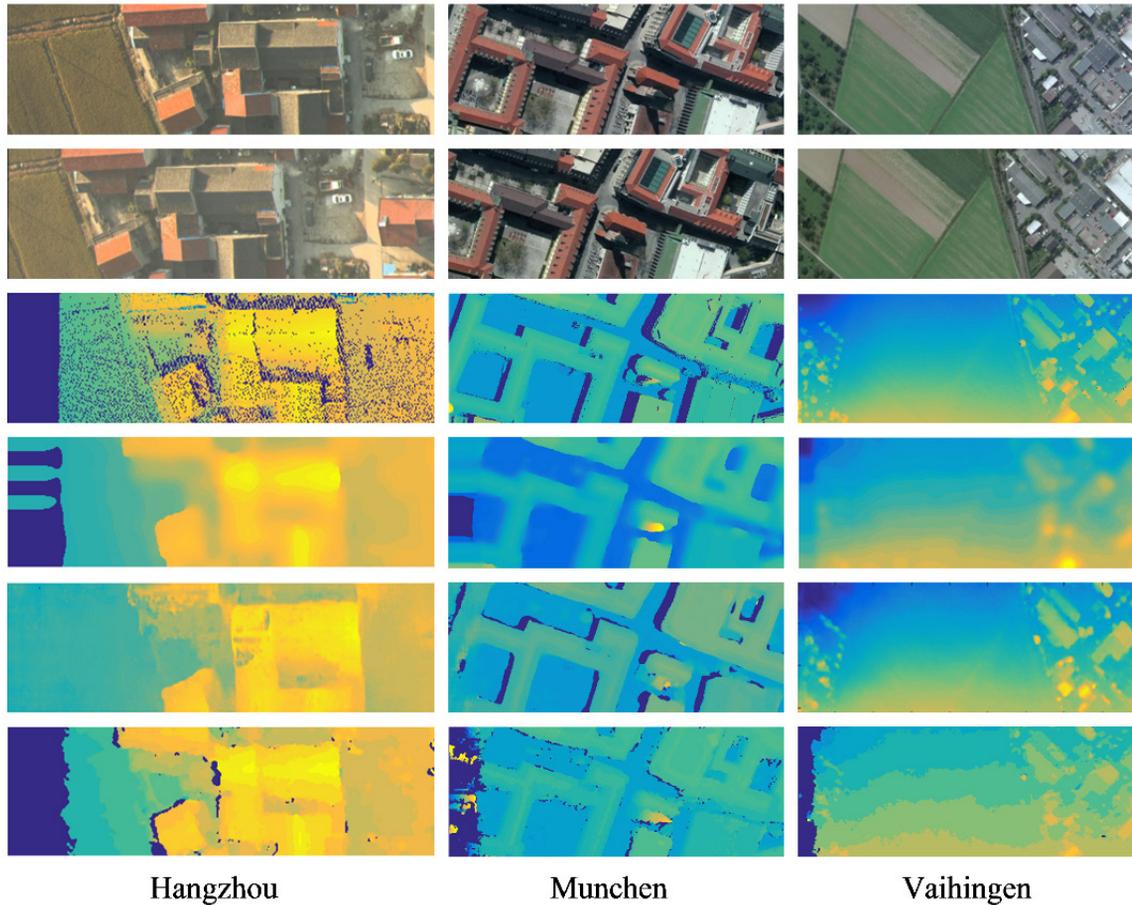


Figure 6. From top to bottom: stereo image pairs, ground truth, and the prediction results of the MC-CNN, GC-Net and SGM methods. For predicting the Hangzhou image with only sparse depth points, the GC-Net model, which requires dense depth points as training samples, is trained on the München data set.

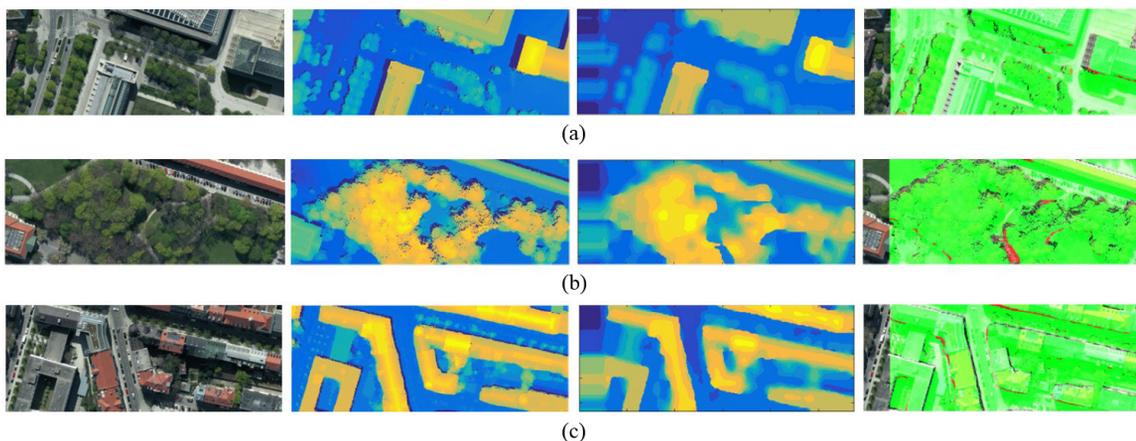


Figure 7. The MC-CNN results on different land cover categories: flat ground (a), trees (b), and buildings (c). From left to right: the left image, ground truth, and the predicted disparity map and residual map. In the residual maps, green indicates disparity error below three pixels, and red indicates disparity error above three pixels.

prediction results of the MC-CNN, GC-Net, and SGM methods. It can be observed that the result of the GC-Net is the closest to the reference map.

Results of applying the MC-CNN method to the München data set are used to analyze the influence of land cover categories. Three image patches, including ground, buildings, and trees, respectively, were selected. The differences between the predicted results and ground truth are shown in Figure 7. Generally, the MC-CNN performs well at all categories and shows no obvious difference. Mismatching appears at the boundary of occluded areas. An obvious mismatching occurred in Figure 7c. It can be observed that the path is partially obscured by tree branches in both or either of the stereo images.

Figure 8 shows the 3D scenes recovered from the dense disparity maps. The result of SURE shows some distortions in the Hangzhou data set. Compared to the other methods, it processes the whole aerial images and is more vulnerable to the accuracy of interior and exterior orientation elements. On the München data set, the results of all the methods are close to the referenced 3D scene and accurate. However, the texture of the buildings' side is much clearer using the SURE, as it utilizes multi-view images, especially those from a side strip. All methods perform well on the Vaihingen data set, which consists of flat landscapes.

Transfer Learning of Deep Learning-Based Stereo Methods

We evaluate direct transfer learning of the MC-CNN and GC-Net (the poorly performed DispNet was omitted). Table 2 shows the test results of the MC-CNN on 3PE and 1PE indicators. The source data set is used for model training and the target data

set for prediction. For example, when using the Hangzhou data itself for training, the 3PE accuracy is 95.3% (bold diagonal elements) on the Hangzhou data set; when using the KITTI2012 for training, the accuracy is 94.4%. The difference between them reflects the degradation of each model using transfer learning (see the last row).

In general, the MC-CNN method has good generalization ability whether the model is trained on close-range images, simulated scenes, or aerial images. The degradation degree varies between 0.2% and 2.2% on 3PE with an average of 0.6% on the three aerial data sets, 0.8% and 5.6% on 1PE, and an average of 2.1% on the aerial data sets. When comparing the 3PE, using a pretrained MC-CNN model is still superior to SGM and equivalent to the SURE software.

Better generalization ability is achieved when the source and target sets are similar. For example, the KITTI2012 model obtained a test accuracy of 95.8% with the KITTI2015, decreasing by only 0.2%. The test accuracy of using the Hangzhou data decreased by 0.5% using the pretrained model on either the München or the Vaihingen data but decreased by 2.2% using the models pretrained on the street-view data. The Vaihingen model shows the worst generalization ability among the three aerial data sets, as the terrain of Vaihingen is relatively flat, and very little information about the dramatic parallax changes is learned.

Table 3 shows the results of direct transfer learning based on the GC-Net model. As only the Driving, München, and Vaihingen data sets contain the dense depth maps, the three data sets are used to train the models, which are then applied to the test data set for prediction.

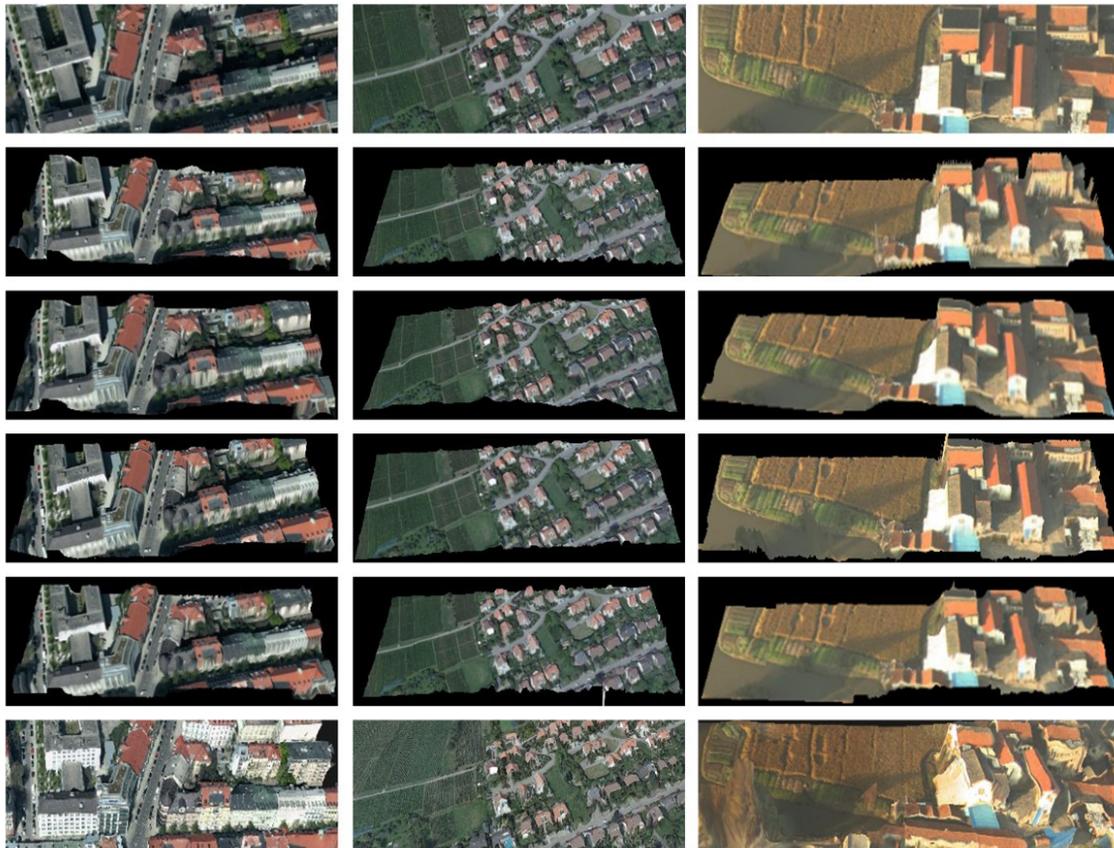


Figure 8. 3D scenes recovered from predicted disparity maps. From top to bottom: the left images, referenced 3D scenes, and the prediction results of the MC-CNN, GC-Net, SGM, and SURE.

Table 2. Pretraining the MC-CNN models on different source data sets and applying them on target data sets.

Target Set	Source Set, Accuracy (3PE/1PE)					
	KITTI2012	KITTI 2015	Hangzhou	München	Vaihingen	Average Degradation
KITTI2012	0.963/0.866	0.957/0.848	0.941/0.856	0.945/0.797	0.946/0.813	-0.016/-0.038
KITTI2015	0.958/0.768	0.960/0.778	0.951/0.761	0.955/0.751	0.953/0.750	-0.006/-0.021
Hangzhou	0.944/0.808	0.942/0.805	0.953/0.816	0.948/0.770	0.940/0.760	-0.010/-0.030
München	0.960/0.854	0.960/0.851	0.960/0.844	0.965/0.867	0.959/0.850	-0.005/-0.017
Vaihingen	0.988/0.919	0.987/0.912	0.987/0.916	0.989/0.922	0.992/0.932	-0.004/-0.015

Table 3. Pretraining the GC-Net models on different source data set and applying them on target data set.

Target Set	Source Set, Accuracy (3PE/1PE)			
	Driving	München	Vaihingen	Average Degradation
Driving	0.926/0.857	0.895/0.808	0.895/0.793	-0.031/-0.057
München	0.969/0.893	0.984/0.953	0.964/0.922	-0.018/-0.046
Vaihingen	0.980/0.881	0.979/0.943	0.997/0.980	-0.018/-0.068
KITTI2015	0.934/0.739	0.881/0.705	0.942/0.743	—/—
Hangzhou	0.911/0.779	0.940/0.799	0.949/0.841	—/—

Table 4. Comparison of stereo and triple-view image matching results on the München data set.

Method	Stereo/Triple-View				
	MC-CNN	GC-Net	DispNet	SURE	SGM
Percentage (< 1 m)	0.922/0.923	0.938/0.938	0.695/0.690	0.902	0.892/0.888
Avg-err (m)	0.590/0.539	0.397/0.392	1.043/1.039	0.635	0.759/0.710

The GC-Net has good generalization ability but slightly worse than the MC-CNN. Compared with the model trained on the same data set, the degradation degree of the model pretrained on other data sets varies between 1.5% and 3% on the 3PE and with an average of 1.8% on the München and Vaihingen data sets, 3.1% and 9.9% on the 1PE, and 5.7% on the two aerial data sets. When transferring between the aerial data sets, the test accuracy decreased by about 2%, while with the MC-CNN, the decrease was only 0.6%.

To sum up, both the pretrained MC-CNN and the GC-Net models can be directly applied to a different data set if the focus is the outlines of 3D scenes with 3PE as an indicator. To identify very fine structures and using the 1PE as an indicator, fine-tuning on small target samples may be required.

Triple-View Geometry for Deep Learning-Based Dense Matching

We evaluate whether using a multi-view geometry could further improve the accuracy of the deep learning-based stereo methods. Table 4 shows the comparison results between stereo and triple-view dense matching on the München data set with two indicators. One indicator is the percentage of the pixels with errors less than 1 m among all pixels (shown as P indicator); the other is the average absolute error of all pixels compared to the georeferenced depth map (shown as A indicator).

Both the stereo MC-CNN and the GC-Net methods have high performance and better than the other methods on the P indicator (Table 4). When triple-view geometry was introduced, the P indicator remains almost the same. It seems that errors larger than 1 m have been mostly eliminated by the postprocessing of the MC-CNN and the integration of the context and geometry information in the GC-Net, respectively. The remaining large errors may be due to the occlusions which cannot be compensated for by the triple-view geometry. Anyway, the parallel is between occlusions and label errors.

In addition, the stereo-based GC-Net performs much better than other methods on the A indicator and reaches 0.397 m. When the multi-geometry is introduced into the model, the GC-Net and MC-CNN improved slightly. The GC-Net outperforms the MC-CNN by 0.137 m, SURE by 0.243 m, and SGM by 0.318 m in depth accuracy, respectively. As the MC-CNN and the SGM

share almost the same structure except the matching cost, they are comparable, and the difference of accuracy (0.171 m) indicates that the matching cost of a simple CNN structure is better than an empirical one. Both show some improvement (0.041 m and 0.049 m, respectively) when the triple-view geometry is introduced. In either case, the DispNet functions the worst.

It could be concluded that when measured by depth instead of parallax, the GC-Net is the best method and that the MC-CNN is slightly better than the SURE. By introducing triple-view geometry, the MC-CNN achieved more significant improvement comparing to the GC-Net.

Figure 9 shows the depth maps of a triple-overlapping area of a whole aerial image generated by the three multi-view methods. It visually demonstrates that the result of the triple-view GC-Net (Figure 9d) is the most similar to the ground truth (Figure 9b). The result of the MC-CNN (Figure 9e) is slightly worse than that of the GC-Net; the result of the SURE shows some holes. From Figure 9c, the holes can be clearly observed. They are caused by the matching failure of the algorithm (or the matching may be not performed pixel-by-pixel) because both the MC-CNN and the GC-Net could find dense and smooth matches in the same area.

Discussion

In this study, we have thoroughly evaluated the performance of deep learning-based dense matching on stereo and multi-view aerial images. To illustrate efficiency, we take the München data set as an example. The overlapped parts of the 14 7072 × 7776 aerial images were almost seamlessly cropped into 300 768 × 384 stereo tiles. It took 5 hours to train the MC-CNN with 80% stereo tile samples; the prediction of a disparity map took 0.6 seconds (about 180 seconds for predicting the whole data set). It took 6.7 hours to train the GC-Net with 80% samples and 0.16 seconds to predict a single disparity map. It took the SURE 4.5 hours to generate the DSM of the aerial images. Thus, the efficiency of the traditional methods and the deep learning-based methods is of the same level. However, if the CNN model has been well pretrained before, the efficiency is much higher than that of a traditional method.

For transfer learning, besides the transfer learning strategy described in the section “Transfer Learning of Deep Learning-Based Stereo Methods”, the other transfer learning strategy conditioned on the target set containing a small number of samples could be considered. Fine-tuning a pretrained model (as initial parameters) with the available samples could improve the performance. Tables 5 and 6 show the results of fine-tuning based on the MC-CNN and GC-Net. DT represents training directly on a target set with random initialized parameters; TL represents model transfer and fine-tuning the parameters with the given samples. The performance of the MC-CNN is evaluated on the Hangzhou data set with a model pretrained on the KITTI2015 (Table 5); the performance of the GC-Net is evaluated on the München data set, with a pretrained model on the Vaihingen (Table 6).

Table 5 shows that an accuracy of 94.3% (79.9%) is obtained when the model is directly trained (DT) with 25 samples. When the sample number is doubled, the test accuracy is improved by only 0.09% (0.012%). This indicates that a large sample size is not necessary to train a satisfactory MC-CNN model. Thus, fine-tuning with a pretrained model (TL) seems to contribute trivially. The test accuracy is 94.9%

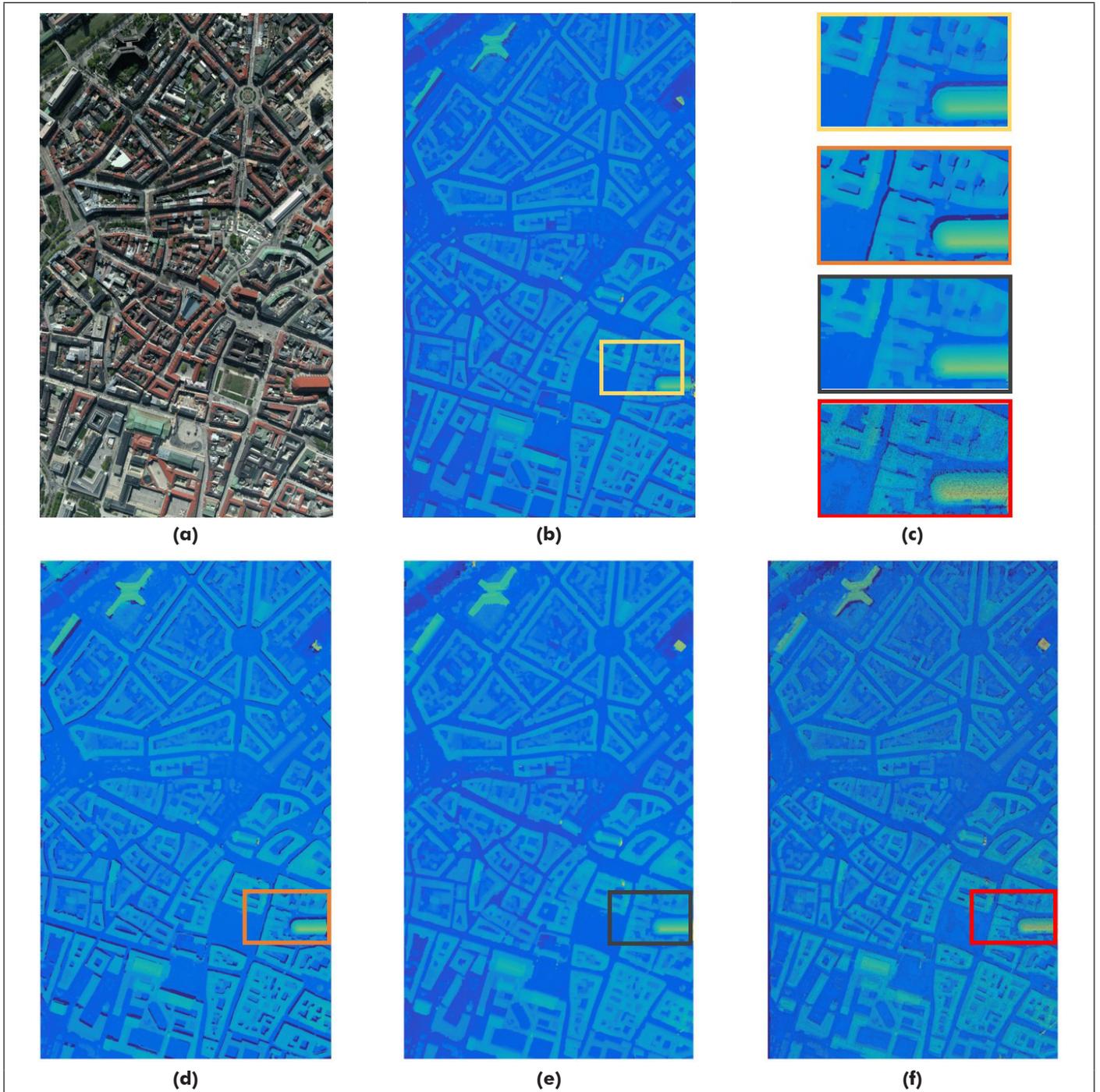


Figure 9. The depth maps of a München image using the multi-view-based methods. (a) a whole aerial image (the triple-overlapping area), (b) ground truth, (d) GC-Net results, (e) MC-CNN results, (f) SURE results and (c) the enlarged parts of the results of the three methods (from top to bottom: ground truth, GC-Net, MC-CNN, and SURE).

in 3PE using 25 training samples for fine-tuning, with a 0.5% improvement compared to the direct training.

In Table 6, when using direct training, the test accuracy is 78.3% (31.6%) with 25 samples; the accuracy reached 90.2% (60.3%) with 50 samples. This indicates that the end-to-end GC-Net requires more training samples than the MC-CNN. When using the pretrained model and fine-tuning strategy, 96.5% (91.6%) accuracy can be achieved with 25 training samples, 18.1% and 60.0% higher than direct training in 3PE and 1PE, respectively.

From the statistical results (Tables 5 and 6), fine-tuning helps improve test accuracy, especially for the end-to-end methods. The fewer the sample number, the greater the

fine-tuning effects. It is also found that fine-tuning can not only improve the accuracy but also reduce the iterations of training.

Conclusion

This study discusses the use of deep learning in the dense matching of aerial images and compares their performance with traditional methods on various data sets, analyzes the generalization ability of deep learning methods, and presents a deep learning-based multi-view dense matching framework. First, in both stereo and triple-view conditions, the end-to-end GC-Net outperforms all the other methods by a large margin. The

MC-CNN, which only learns matching cost, performs slightly better than the SURE. The SGM and the DispNet perform the worst. Second, both the MC-CNN and the GC-Net have shown satisfactory generalization ability, which ensures that a pre-trained model on open training data sets can be directly applied to target aerial images. However, if a high precision is required, a small set of training samples in the target data set could further improve the accuracy through fine-tuning. Finally, for deep learning-based stereo methods, multi-view geometry could further improve the accuracy of the predicted depth map. When available, the use of multi-view information and a pre-trained model as initial parameters could significantly improve the performance of remote sensing dense matching.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (41471288).

References

- Aguilera, C.A., Aguilera, F.J., Sappa, A.D., Aguilera, C. and Toledo, R. Learning cross-spectral similarity measures with deep convolutional neural networks, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2016, pp. 1-9.
- Bleyer, M., Rhemann, C. and Rother, C. PatchMatch Stereo-Stereo Matching with Slanted Support Windows, *Proceedings of the British Machine Vision Conference*, 2011, pp. 1-11.
- Boykov, Y.Y. and Jolly, M.P. Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images, *Proceedings of the IEEE International Conference on Computer Vision*, 2001, pp. 105.
- Chang, J.-R. and Chen, Y.-S. Pyramid Stereo Matching Network, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 5410-5418.
- Dosovitskiy, A., Fischer, P., Ilg, E., Hausser, P., Hazirbas, C., Golkov, V., Van Der Smagt, P., Cremers, D. and Brox, T. FlowNet: Learning optical flow with convolutional networks, *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 2758-2766.
- Furukawa, Y. and Ponce, J., 2010. Accurate, dense, and robust multiview stereopsis, *IEEE transactions on pattern analysis machine intelligence*, 32(8):1362-1376.
- Fusiello, A., Trucco, E. and Verri, A., 2000. A compact algorithm for rectification of stereo pairs, *Machine Vision and Applications*, 12(1):16-22.
- Han, X., Leung, T., Jia, Y., Sukthankar, R. and Berg, A.C. Matchnet: Unifying feature and metric learning for patch-based matching, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3279-3286.
- Hirschmuller, H., 2008. Stereo processing by semiglobal matching and mutual information, *IEEE Transactions on pattern analysis and machine intelligence*, 30(2):328-341.
- Kendall, A., Martirosyan, H. and Dasgupta, S., 2017. End-to-end learning of geometry and context for deep stereo regression, *CoRR*, 04309.
- LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning, *Nature*, 521(7553):436.
- Mayer, N., Ilg, E., Hausser, P., Fischer, P., Cremers, D., Dosovitskiy, A. and Brox, T. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 4040-4048.
- Menze, M. and Geiger, A. Object scene flow for autonomous vehicles, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3061-3070.

Table 5. Fine-tuning results of the MC-CNN on the Hangzhou data set with pretrained model on the KITTI2015 (TL) compared to direct training with the available target samples (DT).

Methods	Samples									
	25 Pairs		50 Pairs		100 Pairs		200 Pairs		300 Pairs	
	DT	TL	DT	TL	DT	TL	DT	TL	DT	TL
Accuracy										
(3PE)	0.943	0.949	0.944	0.948	0.946	0.948	0.951	0.952	0.952	0.953
(1PE)	0.799	0.812	0.801	0.811	0.807	0.812	0.81	0.813	0.811	0.815
Improvement										
(3PE)	0.50%		0.37%		0.14%		0.12%		0.11%	
(1PE)	1.27%		0.97%		0.52%		0.31%		0.35%	

Table 6. Fine-tuning results of the GC-Net on the München data with pretrained model on the Vaihingen (TL) compared to direct training with the available target samples (DT).

Methods	Samples									
	25 Pairs		50 Pairs		100 Pairs		200 Pairs		250 Pairs	
	DT	TL	DT	TL	DT	TL	DT	TL	DT	TL
Accuracy										
(3PE)	0.783	0.965	0.902	0.947	0.928	0.961	0.959	0.977	0.972	0.978
(1PE)	0.316	0.916	0.603	0.899	0.881	0.931	0.904	0.942	0.925	0.944
Improvement										
(3PE)	18.10%		4.50%		3.20%		1.80%		0.60%	
(1PE)	60.00%		29.60%		5.00%		2.80%		1.90%	

- Pan, S.J. and Yang, Q., 2010. A survey on transfer learning, *IEEE Transactions on knowledge data engineering*, 22(10):1345-1359.
- Pang, J., Sun, W., Ren, J.S., Yang, C. and Yan, Q. Cascade Residual Learning: A Two-Stage Convolutional Neural Network for Stereo Matching, *Proceedings of the ICCV Workshops*, 2017.
- Ronneberger, O., Fischer, P. and Brox, T. U-net: Convolutional networks for biomedical image segmentation, *Proceedings of the International Conference on Medical image computing and computer-assisted intervention*, 2015, (Springer), pp. 234-241.
- Rothermel, M., Wenzel, K., Fritsch, D. and Haala, N. SURE: Photogrammetric surface reconstruction from imagery, *Proceedings of the LC3D Workshop*, Berlin, 2012, pp. 2.
- Scharstein, D. and Szeliski, R., 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms, *International Journal of computer vision*, 47(1-3):7-42.
- Schmidhuber, J., 2015. Deep learning in neural networks: An overview, *Neural networks*, 61:85-117.
- Seki, A. and Pollefeys, M. SGM-Nets: Semi-global matching with neural networks, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Honolulu, HI, USA, 2017, pp. 21-26.
- Shaked, A. and Wolf, L. Improved stereo matching with constant highway networks and reflective confidence learning, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4641-4650.
- Zagoruyko, S. and Komodakis, N. Learning to compare image patches via convolutional neural networks, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 4353-4361.
- Zbontar, J. and LeCun, Y. Computing the stereo matching cost with a convolutional neural network, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1592-1599.
- Zhang, K., Lu, J. and Lafruit, G., 2009. Cross-based local stereo matching using orthogonal integral images, *IEEE transactions on circuits and systems for video technology*, 19(7):1073-1079.
- Zhong, Y., Dai, Y. and Li, H., 2017. Self-supervised learning for stereo matching with selfimproving ability, *arXiv preprint arXiv:1709.00930*. *arXiv:1709.00930*.

Semantic Façade Segmentation from Airborne Oblique Images

Yaping Lin, Francesco Nex, and Michael Ying Yang

Abstract

In this paper, oblique airborne images with very high resolution are used to address the problem from aerial views in urban areas. Traditional classification method (i.e., random forests) is compared with state-of-the-art fully convolutional networks (FCNs). Random forests use hand-craft image features including red, green, blue (RGB), scale-invariant feature transform (SIFT), and Texton, and point cloud features consisting of normal vector and planarity extracted from different scales. In contrast, the inputs of FCNs are the RGB bands and the third components of normal vectors. In both cases, three-dimensional (3D) features are projected back into the image space to support the facade interpretation. Fully connected conditional random field (CRF) is finally taken as a post-processing of the FCN to refine the segmentation results. Several tests have been performed and the achieved results show that the models embedding the 3D component outperform the solution using only images. FCNs significantly outperformed random forests, especially for the balcony delineation.

Introduction

Semantic building façade segmentation is an important sub-task for Level of Detail 3 CityGML model generation. These three-dimensional (3D) models are often required in many disciplines, such as urban planning and disaster management. Façade classification's main objective is to distinguish different components on building façades, like roof, window, and balconies. The manual labeling of façade components is time-consuming and not economically affordable and, therefore, an algorithm for the automated façade interpretation could be extremely useful when large urban areas are considered.

In the recent years, machine learning techniques have shown their huge potential in automated interpretation. Traditional classifiers, like random forest [1] and boosting scheme [2], are widely used in object detection [3] and image segmentation [4], using handcrafted features as input. The main drawback of these approaches is the delivery of noisy pixelwise classification, as semantic classes are independently assigned to each image pixel instead of taking advantages of the information provided by surrounding pixel labels.

In this regard, the developments in Convolutional Neural Networks (CNNs) have achieved good image classification, assigning a label to image patches, while fully convolutional networks (FCNs) [5] have more recently allowed the semantic image segmentation, labeling every pixel in an image. Normally, the networks consist of repetitive down-sampling layers and pooling layers. This structure leads to large receptive fields that allow the networks to learn more representative features and make good use of neighboring information at different levels, while their deficiencies are nonsharp boundaries and blob-like shapes in image segmentation results [6].

Noisy labels as well as oversmoothed labeling are problems of traditional classifiers and CNNs, respectively. In this regard,

several researches have taken conditional random fields (CRFs) as a postprocessing to exploit contextual information. CRFs with 4-connectivity and 8-connectivity only capture relative short-range interactions between pixel labels. Noisy labels can be cleaned by limited-range contextual information, but boundaries of small structures can be smoothed. Recently, fully connected CRFs, modeling both local and global spatial dependencies, have shown successful results in the opposite process, disambiguating object boundaries and figuring out tiny structures, especially useful for coarse outputs of FCNs [6].

Looking at the façade classification task, most semantic façade segmentation tasks are performed on dataset from terrestrial views (CMP [7], ECP [8], and eTRIMS [9]). Terrestrial views images have sufficient details on façades, but the acquisition is time-consuming, especially when the task is to achieve 3D city modeling in large urban areas. These acquisitions, then, need to be registered with aerial acquisitions if the final goal is a 3D city model. In comparison, data acquisition from aerial platforms is more feasible for large-area applications. Multi-camera systems give multi-views of urban objects, providing adequate data for large-scale point cloud generation by means of current photogrammetric techniques. Geometrical features from reconstructed point clouds are additional cues in semantic image segmentation which cannot be calculated from single-view images in conventional datasets. Some recent attempts in semantic façade segmentation from multi-view images [10] [11] have been presented, but still they are confined to terrestrial-view images and only [12] explored potentials of aerial images to address the problem. This work is an extension of [12]. [12] simply uses hand-crafted image and point cloud features while this paper utilizes FCNs to learn highly representative features from data for more accurate façade object prediction.

Our work aims to compare the potentials of random forests (and hand-crafted features) with FCNs in semantic façade segmentation task using airborne oblique images as input. Both two-dimensional (2D) and 3D information exploited in these two methods. A fully connected CRF model is implemented to increase segmentation accuracy and improve the visualization in both cases.

This paper is organized as follows: in the section "Related Work" related works in façade interpretation are discussed. In the section "Method" feature extraction for random forest and principles of FCNs and fully connected CRFs are explained. In the sections "Experimental Setup", "Results", and "Discussion" model parameters and experimental results are shown and discussed. Conclusion and possible future work are described in the last section.

The authors are with University of Twente, The Netherlands. Corresponding author: Michael Ying Yang (michael.yang@utwente.nl).

Photogrammetric Engineering & Remote Sensing
Vol. 85, No. 6, June 2019, pp. 425–433.
0099-1112/19/425–433

© 2019 American Society for Photogrammetry
and Remote Sensing
doi: 10.14358/PERS.85.6.425

Related Work

Currently, there are two categories of methods for semantic façade segmentation, namely top-down and bottom-up methods. Top-down methods rely on geometrical grammar to split a single façade into different parts. Bottom-up methods employ multi-class classifiers to assign a label to each image pixel and use postprocessing to optimize segmentation results, like CRF models.

In the top-down paradigm, the façade is recursively separated into smaller façade segments based on image characteristics and division rules. These rules, hierarchically representing the layout of façade objects, are defined based on strong prior knowledge on façade structure or learned from façade dataset. [8] designs six rules to represent the global configuration of façade objects. Pixel-wise labels from random forests are involved in the façade parsing. The limitation is that their rules are defined for Haussmannian-style buildings in Paris and hardly fit other architectural styles. With the intention of relieving the strong dependency on prior knowledge, [13] uses a Bayesian Model Merging to learn shape grammar for a certain architectural style from labeled façades. However, their approach can only deal with grid-shape façade objects which are well aligned (horizontal lines along rows and vertical lines along columns of the image) and cannot solve the façade segmentation in airborne images where façades are always randomly oriented.

Bottom-up methods get rid of prior knowledge on façade layout. The semantic segmentation is achieved by using machine learning classifiers to assign labels to pixels or superpixels. [14] uses Textonboost to label each façade pixel. Outputs are very noisy, due to contextual information deficiency. [15] adapts a structured random forest to façade interpretation, producing noisy free segmentation. [16] uses the FCN to address the problem. A symmetry loss is added to a convolutional loss function because man-made façade objects are always regular in shape.

Conditional random fields are commonly used as a post-processing to denoise pixel-wise classification results. A hierarchical CRF is proposed by [16] which consists of three terms. The unary term is the probability distribution from a Random Forest classifier. The pairwise term is a color contrastive Potts model exploiting label compatibility between neighboring pixels. The hierarchical term uses mean shift superpixels derived at different scales to exploit the spatial dependencies of façade objects from local to global. With the reduced computational complexity in the fully connected CRF model [17], [14] implements it to semantic façade segmentation. Textonboost is picked to get unary potentials and a linear combination of Gaussian kernels chosen as the pairwise potential to connect every pixel over the whole image. This fully connected structure not only performs well in enforcing the label consistency among nearby pixels, but also detects small façade elements and delineates crisp boundaries. A three-layered approach is designed by [18] to address façade interpretation. The first layer is a label probability distribution of superpixels which is obtained by a trained recurrent neural network. The second layer consists of window and door probability maps computed from object detectors. The first two layers are combined in a CRF model. Weak architectural rules are added to the top layer to structure façade layouts.

Only image features are used in all above works to achieve semantic façade segmentation. In fact, 3D data also benefits urban scene interpretation. [19] improves informal settlement classification by combing predefined 2D and 3D features. [20] adds hand-crafted 3D features to image features learned from CNNs for building damage detection from very high resolution oblique airborne images. The involvement of point cloud features contributes to 3% improvement in average classification

accuracy [20]. Both [19] and [20] involve hand-crafted features which rely on prior knowledge of the dataset, while we feed both 2D and 3D information together into an FCN to allow the learning of highly representative features. Also, [12] adds 3D features to 2D image features in a random forest classifier for semantic façade segmentation, which gives rise to over 20% increase in overall accuracy. However, it only performs three-class classification (roof, wall, and window) which leaves some façade objects whose delineations could benefit from 3D geometries, like a balcony. [10] simply concatenates image and point cloud features as a feature vector for each pixel. Although contextual cues are given in vectors, each vector is independently fed to an ensemble of classifiers. The lack of concurrency in pixel prediction misses global optimality. [21] combines 2D and 3D information at super-pixel level to achieve semantic segmentation of indoor RGB-Depth images. Comparing to pixel-wise labelling in our work, unsupervised segmentation in the first step may lead to inaccurate boundaries that are difficult to be corrected in the following steps.

Not constrained to image pixel labeling, point cloud labelling in 3D space also benefits from 2D and 3D feature integration. Both [22] and [23] extract spectral information from aerial images for each point and a CNN-based method is designed to improve semantic segmentation of noncolor airborne laser scanning point clouds. However, as a result of occlusion, limited laser scanning points are available on vertical surfaces, like facades. Irregular spacing and large sparsity make difficulties in extracting representative features to achieve accurate labelling. [11] conducts 3D labeling on dense matching point cloud through an end-to-end 3D pipeline, integrating image features with point cloud features, including normal vectors, depths, heights, and spin image descriptors at different scales, in a Random Forest classifier. A CRF model which connects four nearest surrounding points is then used to smooth results. This work shows how the integration of both 2D and 3D features and the use of superpixels and object detectors can assure better accuracies.

Currently, most studies focus on façade segmentation in single view images and very few studies exploit potentials of 3D information computed from multi-view images. Many efforts have been spent in scene understanding in aerial images but studies reaching façade level are quite rare. This work investigates potentials of airborne images to address the problem by FCN combined with a fully connected CRF.

Method

Feature Extraction

In this work, image features and point cloud features are extracted for each façade. Then point cloud features are projected back to image space.

2D Feature Extraction

Three types of features are used in our work:

Color features. In our work, spectral information is three features presented in RGB color space.

SIFT. Scale-invariant feature transform (SIFT) descriptor consists of 128 features. For a certain pixel, it is a gradient histogram that is calculated based on gradient orientations and magnitudes at eight fixed orientations over a small image region whose center is that pixel [24] [25].

LM filter. 48 texture features are produced by Leung-Malik filter bank which is composed of Gaussian kernels, Laplacian of Gaussian kernels and other derivatives of Gaussian kernels at different scales [26].

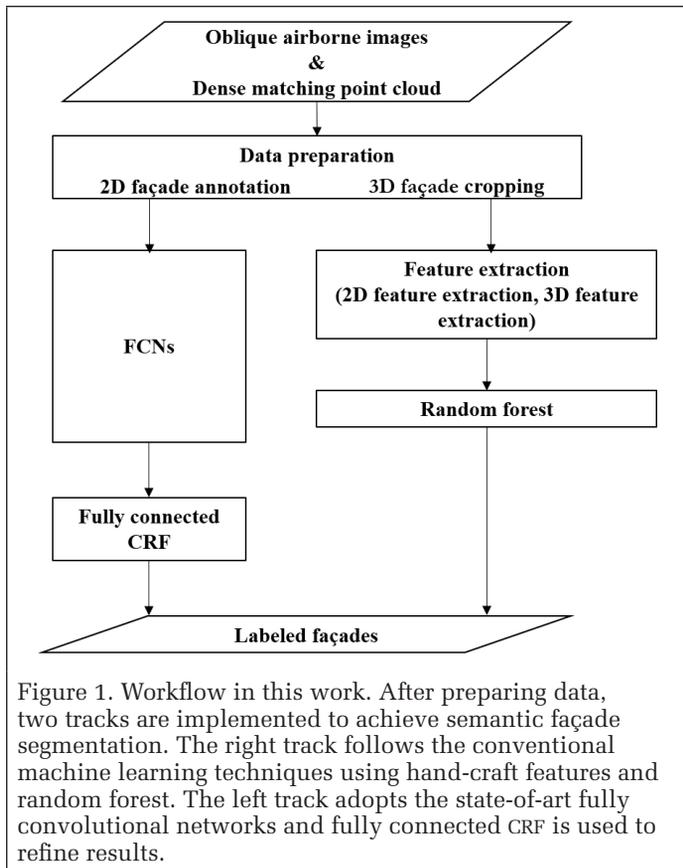


Figure 1. Workflow in this work. After preparing data, two tracks are implemented to achieve semantic façade segmentation. The right track follows the conventional machine learning techniques using hand-craft features and random forest. The left track adopts the state-of-art fully convolutional networks and fully connected CRF is used to refine results.

3D Feature Extraction

3D features used in this paper are normal vectors and planarity which are computed based on neighboring points. Normal vector is useful to differentiate points lying on different surfaces. Planarity is a good indicator to infer the flatness of a surface and distinguish objects with different forms [27]. Planarity is calculated from normalized eigenvalues ($e_1 + e_2 + e_3 = 1$) of the covariance matrix that is derived from 3D coordinates of defined neighbors.

$$Planarity = \frac{e_2 - e_3}{e_1} \quad (1)$$

The strategy and the range of the searching are two key elements when selecting local neighbors around a point. “K-nearest neighbors” is the searching strategy in our work. In terms of the searching range, instead of extracting features at a single scale, our work computes both normal vector and planarity from 20, 100, and 500 nearest points respectively. This is because features extracted from a single scale are inadequate to describe objects and the variation of 3D features at multi-scales can be a signature for small objects and flat surfaces [28]. Objects on planes can be detected, like vertical balcony surfaces on walls [12].

Feature Combination

The integration of 2D and 3D features is achieved by projecting 3D features back into oblique airborne images based on their Pmatrix. The Pmatrix is produced by Pix4D software during the point cloud generation from oblique images. 3D features can be related to image patches with different size. Pixels in the same patch share the same 3D features. If more than one points fall in the same image patch, corresponding features are averaged to assign values to that patch. In practice, if the patch is too small, many voids are left in image space. On the

contrary, if the patch is too large, the void percentage will decrease but the projected 3D features will be too coarse to give detailed information, averaging the information. As full resolution point cloud was not used in our work to keep the balance between void percentage and details of information, 4×4 is picked as the optimal patch size during the projection.

Random Forest

Random forest is composed of many independent decision trees and classification results is a histogram accumulated on those trees. Each decision tree is a classification function of n features to get a probability distribution over a label space. Features of a sample are recursively classified by branching down the tree to a leaf node. For each node in the tree, a split function is learned to decide the path of a sample to reach a leaf node according to the values of the sample features. The splitting terminates at a leaf node where the sample is assigned to a class label.

Fully Convolutional Neural Network

Typical deep convolutional neural networks are made up of a sequence of layers. A convolutional layer is usually followed by nonlinear activation layers that bring nonlinearity to the network and therefore allow the network to learn more complicated and representative features. A pooling layer is always set on the top of an activation layer, summarizing the filter responses to downsample the feature map and thus learn features at a higher level.

FCNs replace the fully connected layers in typical classification networks with fully convolutional layers [5]. The size of the response map of the last convolutional layer is always smaller than the initial image due to the downsampling effect in previous layers. Therefore, the shrunken feature map should be upsampled back to the initial size by a deconvolutional layer.

In our work, vgg16 is modified for semantic segmentation according to the strategy in [5]. All layers are kept as layers in the base network, except last two fully connected layers replaced by two fully convolutional layers which are initialized by random number. Bilinear filters are added to upsample the final feature map back to the initial image size. If the final feature map is directly upsampled to the initial size, the prediction is likely to be coarse and inaccurate [5]. In vgg16, the final feature map is supposed to enlarge by 32 times. Therefore, feature maps from former layers are also concatenated to the final feature map in the final pixel-wise prediction phase [5]. In the FCN (Figure 2), the final feature map is firstly upsampled by twice and then combined with the feature map of pool4. The concatenated feature is enlarged by 16 times to give pixelwise prediction (FCN-16s). Next, the integrated feature map is upsampled twice and then combined with the feature map of pool3. After this, the new feature map only needs to be upsampled by eight times back to the initial size (FCN-8s). This could give finer and more accurate semantic segmentation results.

Conditional Random Field

Conditional random fields are commonly used as a post-processing of machine learning classifiers to denoise semantic segmentation results. In 4-connectivity and 8-connectivity CRF models, only short-range spatial dependencies are built to involve limited contextual information. In fully connected CRFs, every pixel connects to the rest of the pixels over the image and this fully connected neighboring system allows the modeling of longer spatial interactions. Graph-cut is a conventional method to solve CRF models with simple connections in image segmentation [29]. However, it is not applicable to fully connected CRF due to the computational complexity. Therefore, [17] uses a linear combination of Gaussian kernels to represent the pairwise term and the mean field approximation

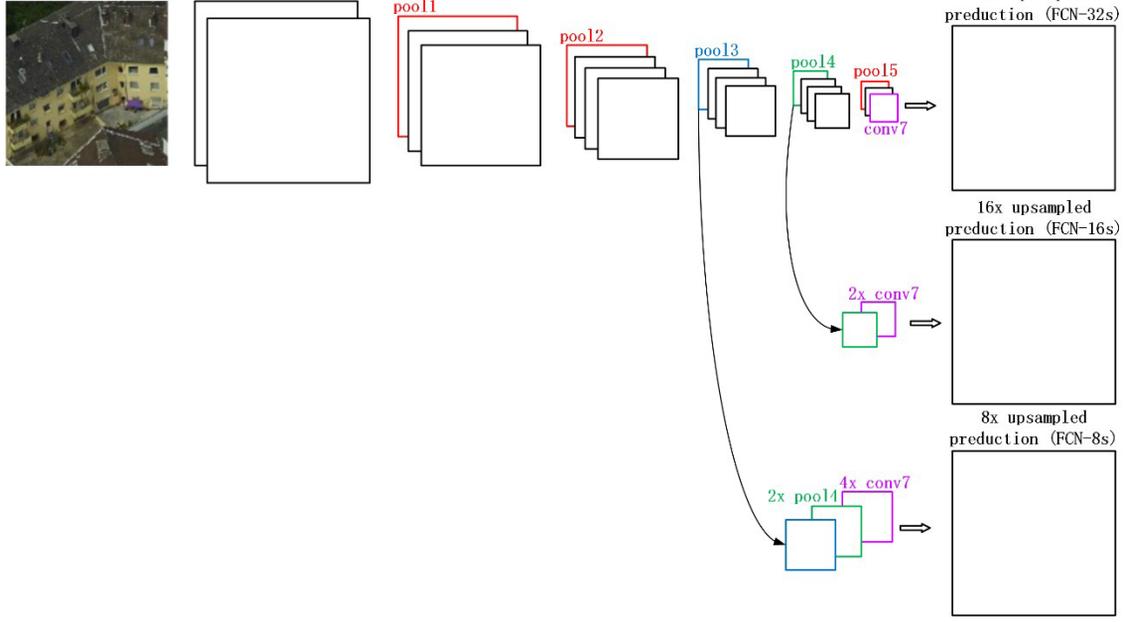


Figure 2. FCN structure derived from the conventional vgg16 net to combine coarse and fine feature maps for pixelwise prediction. The size of pool1, pool2, pool3, pool4, and pool5 are 1/2, 1/4, 1/8, 1/16, and 1/32 of the initial image size respectively.

to give an interference of fully connected CRF. In our work, a set of random variables $\{x_1, \dots, x_N\}$ is constructed to form a random field \mathbf{X} , where N denotes the number of pixels in a whole image. The domain of a random variable is a label set $\mathcal{L} = \{l_1, \dots, l_k\}$, where k represents the number of classes. The random field \mathbf{X} is conditioned on a set of image features $\mathbf{I} = \{I_1, \dots, I_N\}$. This conditional random field (\mathbf{I}, \mathbf{X}) is a Gibbs distribution and it is written as:

$$P(\mathbf{X}|\mathbf{I}) = \frac{1}{Z(\mathbf{I})} \exp\left(-\sum_{c \in \mathcal{C}_G} \phi_c(\mathbf{x}_c | \mathbf{I})\right) \quad (2)$$

$$E(\mathbf{x}) = \sum_{c \in \mathcal{C}_G} \psi_c(\mathbf{x}_c) \quad (3)$$

$$Z(\mathbf{I}) = \sum_{\mathbf{x}} \exp(-E(\mathbf{x})) \quad (4)$$

Here, an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is built over \mathbf{X} . $\phi_c(\mathbf{x}_c | \mathbf{I})$ is a potential function of all variables ($\mathbf{x}_c = \{x_i, i \in c\}$) in a clique c . The collection of all cliques over the graph \mathcal{G} is denoted by \mathcal{C}_G . $E(\mathbf{x})$ is a sum of all potentials and this Gibbs energy function aims to label random variables $\mathbf{x} \in \mathcal{L}^N$. $\psi_c(\mathbf{x}_c)$ is a simplified expression of $\phi_c(\mathbf{x}_c | \mathbf{I})$. $Z(\mathbf{I})$ is a normalization constant which acts as partition function. The maximum a posteriori labeling \mathbf{x}^* is expressed as:

$$\mathbf{x}^* = \operatorname{argmax}_{\mathbf{x} \in \mathcal{L}^N} P(\mathbf{X}|\mathbf{I}) \quad (5)$$

Where the labeling \mathbf{x}^* is optimized by minimizing the energy function $E(\mathbf{x})$.

Fully connected CRF is composed of two terms and the energy function is written as:

$$E(\mathbf{x}) = \sum_i \psi_u(x_i) + \sum_{i < j} \psi_p(x_i, x_j) \quad (6)$$

where $\psi_u(x_i)$ is the unary potential representing the cost of pixel i to take label l_i , calculated by a classifier (i.e., Random Forest or FCN). The binary potential encourage consistency in pixels which are close in position and have similar image features.

Unary potentials: The probability distribution of x_i over label set $\mathcal{L}(x_i | \mathbf{I})$ is computed by random forest (section "Random Forest") and FCN (section "Fully Convolutional Neural Network") according to both image and point cloud features. The unary potential for x_i is written as:

$$\psi_u(x_i) = -\log P(x_i | \mathbf{I}) \quad (7)$$

Pairwise potentials: The fully connected pairwise term consists of a linear combination of Gaussian kernels [17], shown as:

$$\psi_p(x_i, x_j) = \mu(x_i, x_j) \sum_m w^{(m)} k^{(m)}(\mathbf{f}_i, \mathbf{f}_j) \quad (8)$$

$$\mu(x_i, x_j) = \begin{cases} 0 & \text{if } x_i = x_j, \\ 1 & \text{otherwise,} \end{cases} \quad (9)$$

$$w^{(1)} k^{(1)}(\mathbf{f}_i, \mathbf{f}_j) = w^{(1)} \exp\left(-\frac{|p_i - p_j|^2}{2\theta_\alpha^2} - \frac{|I_i - I_j|^2}{2\theta_\beta^2}\right) \quad (10)$$

$$w^{(2)} k^{(2)}(\mathbf{f}_i, \mathbf{f}_j) = w^{(2)} \exp\left(-\frac{|p_i - p_j|^2}{2\theta_\gamma^2}\right) \quad (11)$$

Where $\mu(x_i, y_j)$ is a typical Potts model that assigns penalties when i and j are different in labels. $k^{(m)}$ denotes Gaussian kernels and $w^{(m)}$ are the corresponding weights. This work uses contrast-sensitive two-kernel potentials: f_i and f_j are two feature vectors of two neighboring pixels i and j , taking both position (p_i and p_j) and color information (I_i and I_j) in Equations 10 and 11. $k^{(1)}(f_i, f_j)$ is an appearance kernel that forces pixels that share similar colors and are close in position to belong to the same class. That means when feature vectors f_i and f_j are similar but pixel i and j are not in a same class, a high penalty will be assigned to encourage the labeling configuration move to another stage that achieves better coherency between pixels. θ_α and θ_β determine penalty values through controlling the similarity degrees of position and color. $k^{(2)}(f_i, f_j)$ is a smoothness kernel that aims to remove small and isolated components.

The inference of this fully connected CRF follows the method designed by [17] based on mean field approximation.

Experimental Setup

Airborne images used in this paper were acquired by an IGI Pentacam system over the city of Dortmund (Germany) on 7 July 2016. Average ground sampling distance for oblique images is 4.5 cm. Pix4D oriented these high-resolution images to produce dense matching point cloud at an urban scale.

In our work, four building components are identified: roof, wall, opening, and balcony. Roof and wall are structures covering a building horizontally and vertically, respectively. Opening refers to structures that allow the passage of light, sound, and air, including windows and doors. “Balcony” is defined as a small platform that protrudes from or intrudes into the wall surface and is accessed by an opening. Since we are interested in building areas, façades are manually cropped from aerial images at first. Then corresponding façade point clouds are manually cropped from large-scale dense matching point clouds. Online annotation tool LabelMe is used to prepare ground truths of façades in image space. Our dataset consists of 250 façades: 160, 35, and 55 façades are used for training, validation, and testing respectively.

Random Forest

In our work, 50 trees are chosen as a trade-off between accuracy and training time. The minimum leaf size that controls the depth of the decision tree is set to be 50, with the intention to avoid overfitting. For each node, 14 features are randomly picked to keep the balance between strength of an individual tree and correlation between different trees [30].

Fully Convolutional Network

In our work, FCN was implemented in MatConvNet framework. Due to the limited training dataset, we fine-tuned a

pretrained network for semantic façade segmentation. As mentioned in the section “Random Forest”, a vgg-16 network trained on ImageNet dataset was downloaded from MatConvNet website. Due to the limited graphics processing unit space, façade images were cropped into 224 × 224 patches to feed into the FCN. For the network only using 2D information, RGB channels were directly put into the FCN. With the purpose of including 3D information, the projected third components of normal vectors were added to RGB images as the fourth channel. Normal vectors were computed from nearest 100 neighboring points. As the downloaded network only allows three-bands inputs to the network, to feed the fourth channel into the network, the filter dimension of the first convolutional layer in vggnet was modified from $3 \times 3 \times 3 \times 64$ to $3 \times 3 \times 4 \times 64$. The added weights were initialized by random numbers. The 2D network was trained for 15 epochs and the 3D network was trained for 20 epochs. Both of them implemented with a dropout rate of 0.5 and a learning rate of 0.001. The momentum and weight decay for both networks were 0.9 and 0.0005, respectively. During the training, flipping of patches and PCA color augmentation [31] were performed for data augmentation purpose. During the testing stage, every testing façade image was cropped into 224 × 224 patches with 50 pixels overlap in both vertical and horizontal directions. Then, patches were semantically segmented by FCNs and labeled patches were concatenated back to the initial façade image.

Fully Connected CRF

Parameter configuration tuned by 35 validation façades is shown as below:

$$w^{(1)} = 1, \theta_\alpha = 3, \theta_\beta = 10, w^{(2)} = 1, \theta_\gamma = 2$$

In our case, the optimal spatial standard deviation θ_α is 3 pixels and the optimal value for color standard deviation θ_β is 10. The influence of θ_α and θ_β on overall pixel accuracy are assessed qualitatively (Figure 4) and quantitatively (Figure 5). For this assessment, $w^{(1)}$ is kept as 1 and $w^{(2)}$ is set to be 0. The accuracy variation is complex with changing θ_α and θ_β but it is obvious that long-range connections cause some failures (Figure 5). In contrast to what is mentioned in [17], most of the spatial standard deviations larger than 35 pixels, relatively short-range connections are more suitable for façade interpretation from aerial oblique images.

Accuracy Assessment

Three measures were used to evaluate semantic segmentation accuracy in different schemes, namely, overall pixel accuracy, averaged pixel accuracy for each class and the average of intersection over union (IoU) for each class [32]. These three measures are calculated in terms of true positives (TP), false



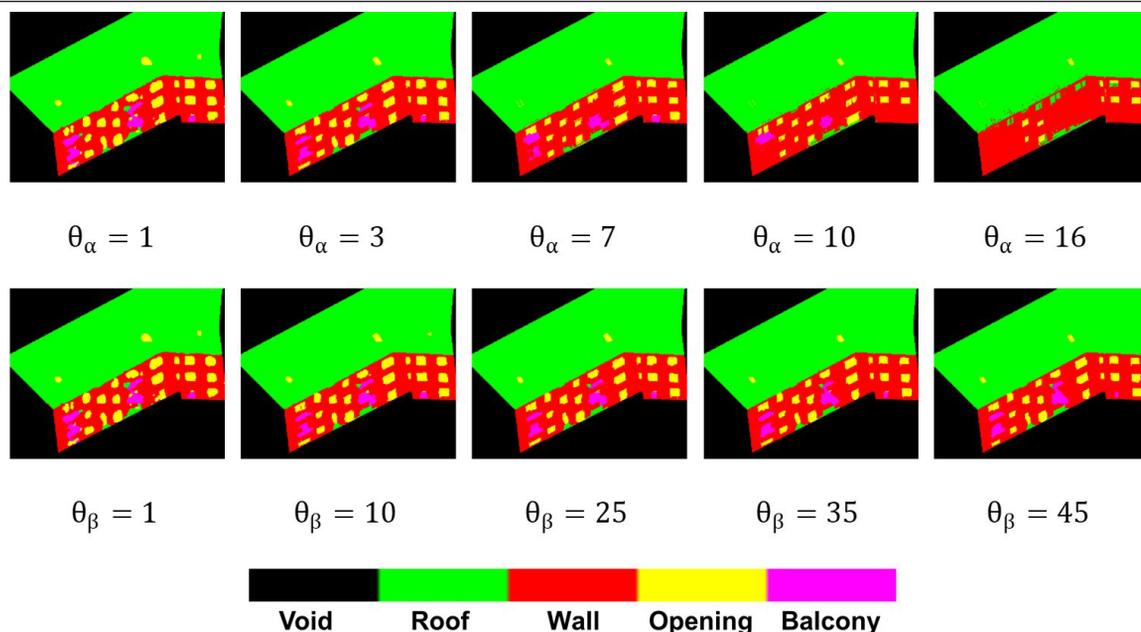


Figure 4. Qualitative assessment of the influence of connections in fully connected CRF (ground truth refers to Figure 3).

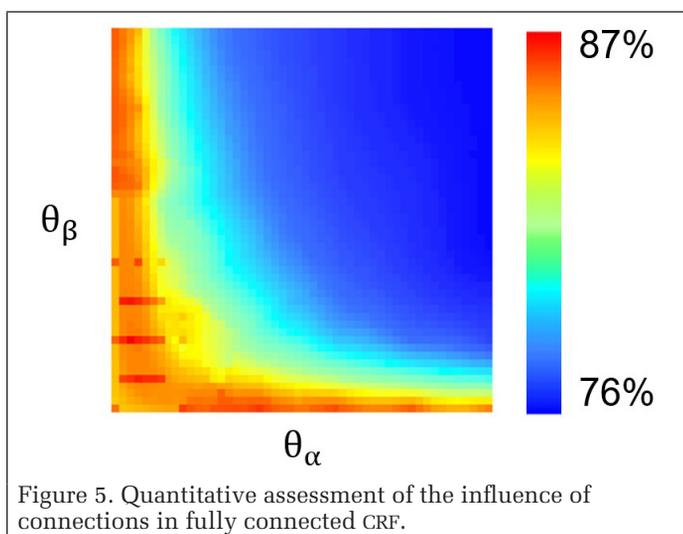


Figure 5. Quantitative assessment of the influence of connections in fully connected CRF.

positives (FP) and false negatives (FN). Overall accuracy is defined as $TP/(TP+FN)$, which is calculated over the whole image. Average accuracy is calculated for every class and then averaged. IoU score is defined as $TP/(TP+FN+FP)$, which is calculated for every class and then averaged.

Results

55 façades were used to test five models. Semantic segmentation results in different schemes are shown in Table 1.

Performance of the random forest using only hand-craft 2D features was the worst in terms of three accuracy measures. Although 90.27% of roof pixels were correctly labeled, accuracies of wall and opening were 59.97% and 38.55%, respectively. Openings on roofs were hard to be labeled (Figure 6) and the classifier was not able to label balcony pixels providing 0.4% accuracy. Figure 6 illustrates that balcony pixels were likely to be labeled as roof and wall.

Adding multi-scale 3D features to a 2D random forest classifier, the overall pixel accuracy and IoU improved by 11% and 11.8%, respectively. Except on roof, accuracies of all other

Table 1. Quantitative results got from five models (55 façades for testing).

Class	RF2D (%)	RF3D (%)	vgg2D (%)	vgg3D (%)	vgg3DCRF (%)
Roof	90.27	87.84	93.42	96.10	90.02
Wall	59.97	88.81	77.03	81.81	81.52
Opening	38.55	53.12	63.90	70.54	84.65
Balcony	0.40	2.05	41.85	31.33	53.83
Average class accuracy	47.30	57.95	69.05	69.95	77.51
Overall pixel accuracy	69.01	80.01	81.63	85.16	85.66
IoU	36.49	48.29	56.21	60.12	60.40

* RF2D represents the random forest trained by 2D features. RF3D represents the random forest trained by both 2D and 3D features. Vgg2D represents the vgg-16 net fine-tuned by RGB images. Vgg3D represents the vgg-16 net fine-tuned by both RGB images and 3D features. vgg3DCRF is a fully connected CRF model using outputs from Vgg3D as the unary term.

classes were improved. Figure 6 demonstrates that 3D features converted most of roof pixels on vertical surfaces to wall or opening pixels and also, turned wall pixels on horizontal surfaces to roof pixels, while confusions between wall and openings remained. Still only 2.05% of the balcony pixels were correctly and some noisy labels can be found in Figure 6 which deteriorate the segmentation results.

The FCN only exploiting RGB channels achieved 81.63% in overall pixel accuracy and 56.21% in IoU. It performed even better than the Random Forest taking both 2D and 3D hand-craft features. Its IoU was 7.92% higher than the best random forest classifier results. Most importantly, the biggest improvement was that 41.85% of the balcony pixels were correctly labeled. Also, few small openings on roofs were identified. Unlike the noisy results got from random forest classifiers, the FCN was not good at delineating object boundaries and produced oversmoothed results.

By involving the third component of normal vector, vgg3D outperformed vgg2D by 3.53% in overall accuracy and 3.91% in IoU. This improvement was much less than the improvement brought to random forest classifiers. Figure 6 illustrates that confusions between roof and wall were largely solved. Opening results were more satisfying, but balcony classification was less successful than in the vgg2D experiment. This is probably due to the quality of the point cloud that is not able to reconstruct in an accurate way small objects on a building façade.

Fully connected CRF gave little contributions to labeling results: it improved 0.5% in overall pixel accuracy and 0.28% in IoU. However, looking at the single classes, pixel accuracies of opening and balcony were significantly improved as visible in the last column in Figure 6 where boundaries of opening and balconies as well as balconies were straight and sharp.

Discussion

Random Forest and Fully Convolutional Network

Random forest has proven to be effective for semantic façade segmentation [16]. Although selected features like color, SIFT and Texton were able to distinguish different classes in terrestrial images (e.g. eTRIMS), these features resulted insufficient for an airborne oblique dataset capturing several architectural styles. FCN [5], a deep learning based approach, is one of the most popular methods for semantic segmentation. The outperformance of FCNs suggested that features learned from the dataset were more representative in our case. These complex representations were able to tell the differences between roofs and balconies too. This result was not possible using hand-craft features in accordance with the achievements presented in [33].

Segmentation results from random forest classifiers were quite noisy because pixel labels were predicted independently without taking contextual information. FCN-8s concatenates feature maps from previous layers to solve the coarse segmentation results. However, over-smoothed boundaries still existed in the presented results, as a consequence of patch downsampling. This suggested that post-processing, like CRFs, was necessary to refine results.

3D Features

In both random forests and FCNs, 3D features showed good performances in solving confusions between pixels on different surfaces. One reason was that normal vector could efficiently separate roof and balcony pixels from wall and opening pixels. Higher improvements were achieved in the Random Forest than in the FCN-8s. This suggests that 3D features are more important when 2D features are not able to provide efficient representations to classes on different surfaces. However, confusions between wall and opening pixels, due to the deficiency of 2D features, were still not solved. The 3D information cannot contribute in the labeling of classes with similar geometries (at the resolution of the used images). In this regard, inaccurate

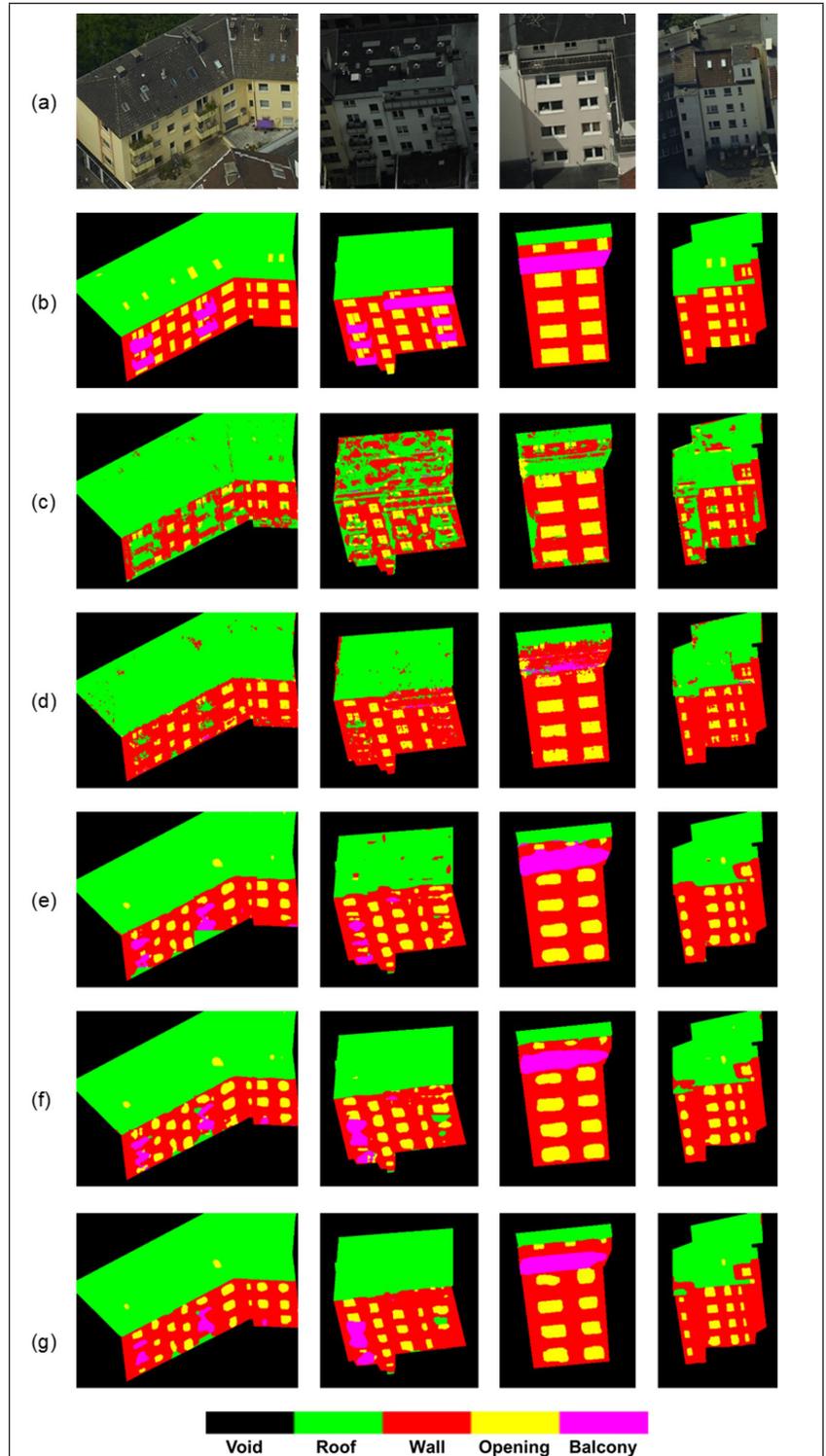


Figure 6. Examples from our dataset. (a) cropped façade images from oblique aerial images. (b) ground truths. (c) results from random forest using 2D features. (d) results from random forest using 2D and 3D features. (e) results from the vgg16 net fine-tuned by RGB images. (f) results from the vgg-16 net fine-tuned by both RGB images and 3D features. (g) results achieved refining the results of (f) with fully connected CRF.

point clouds produced by poor image dense matching could even weaken the ability of 3D features to solve misclassifications. This is clearly shown in Figure 7 where FCN-8s using only 2D features could roughly delineate balcony boundaries while these boundaries were almost ignored introducing the 3D information given by the third components of normal vectors. As depicted in the same figure, very few points are in correspondence on balconies and computed normal vector was not able to define differences between the balconies and the wall. These misclassifications could hardly be corrected by CRF models.

Fully Connected CRF

Taking fully connected CRF as the postprocessing of FCN-8s only gave a little improvement in accuracy, but it refined results and showed better visualization (Figure 6). Compared with [17], a small value of θ_a was adopted. This suggests that very long-range interactions are not helping the object recognition and segmentation in the presented application. As mentioned in [17], most of θ_a are larger than 35 pixels but relatively short-range connections are more suitable for façade interpretation from aerial oblique images. In this study, the θ_a value was 3 pixels in our work. This value of short-range contextual information could be explained by regular shape of man-made façade objects.

Conclusion

This paper gave an investigation of semantic façade segmentation from airborne images. Four classes (roof, wall, window, and balcony) were identified in this work. The problem was addressed by random forests and FCNs. Results suggested that FCNs which learned features from dataset performed much better than random forests which used hand-crafted features. This is in agreement with many other works dealing with similar applications. In this work, FCN using 2D features got 19.72% higher IOU than the random forest taking RGB, SIFT, LM filter bank features. To refine the segmentation results,

fully connected CRF was implemented as a postprocessing step of the FCN results, exploiting both 2D and 3D information. It helped to delineate more accurate object boundaries, but it only contributed to little improvements in accuracy.

We preform semantic segmentation from manually cropped façades while automatic identification of the buildings is still an open research question. Fast and accurate recognition of buildings from large scale data in both 2D and 3D space leaves to be solved. For the future work, more classes could be identified in façade segmentation. The rectification of the oriented façades could be a future option to simplify the classification task. As objects on façades always have regular shapes, soft constraints could be added to regulate segmentation results too. More recent advanced network structures will be tested and further customized for this specific task in the future. Also, this research relies on image segmentation and 3D descriptors are projected to 2D space, while the perspective work could apply innovative network architectures to label points in 3D space.

Acknowledgments

The authors would like to express their sincere gratitude to IGI mbH and Aerowest GmbH companies for providing the aerial image dataset over Dortmund city center.

References

- [1] Frohlich, B., E. Rodner and J. Denzler. A fast approach for pixelwise labeling of facade images. Pages 3029–3032 in *2010 20th International Conference on Pattern Recognition*, 2010.
- [2] Shotton, J., J. Winn, C. Rother and A. Criminisi. TextonBoost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. Pages 1–15 in *European Conference on Computer Vision*, 2006.

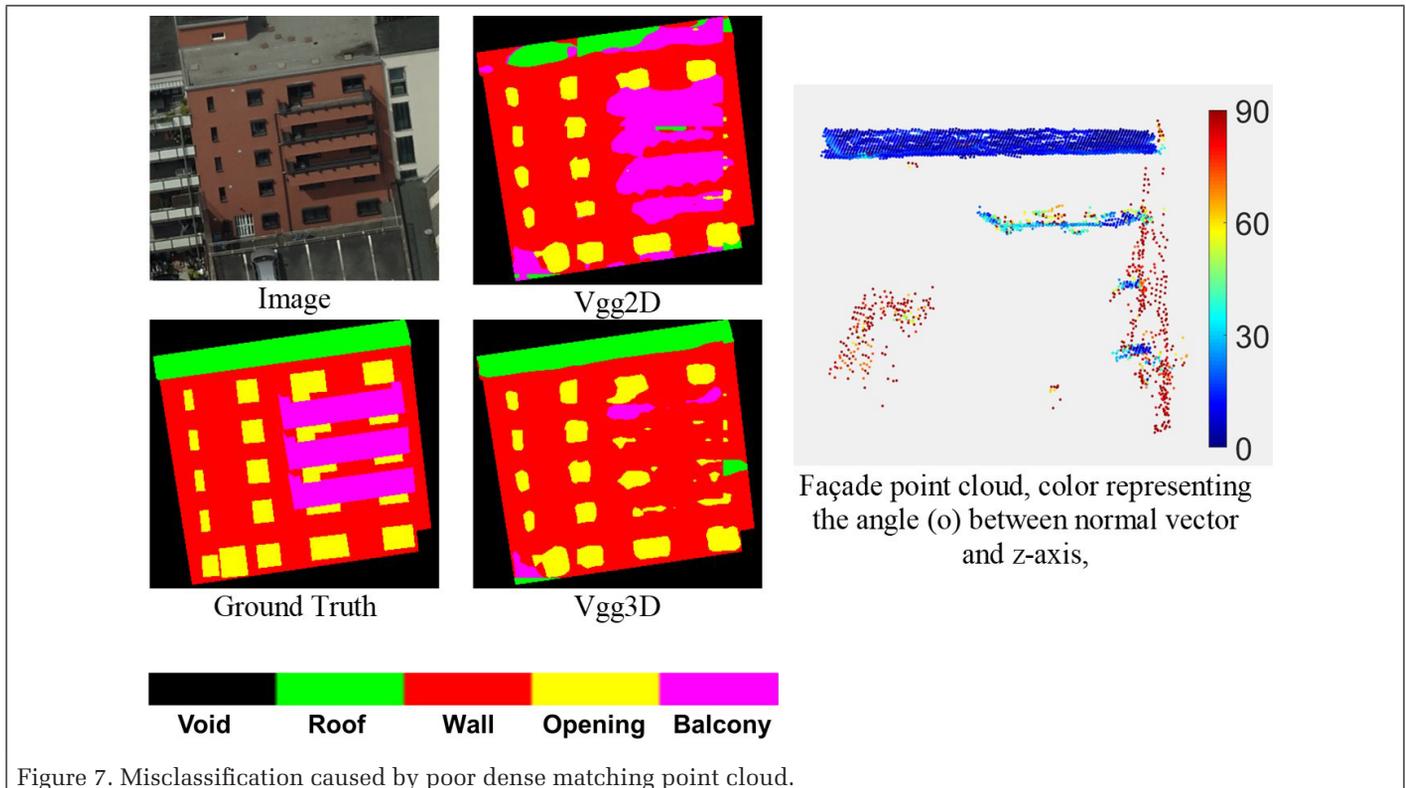


Figure 7. Misclassification caused by poor dense matching point cloud.

- [3] Torralba, A., K. P. Murphy and W. T. Freeman. Sharing features: Efficient boosting procedures for multiclass object detection. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2004.
- [4] Shotton, J., J. Winn, C. Rother and A. Criminisi. 2009. TextonBoost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context. *International Journal of Computer Vision (IJCV)*.
- [5] Long, J., E. Shelhamer and T. Darrell. Fully convolutional networks for semantic segmentation. Pages 3431–3440 in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015.
- [6] Chen, L.-C., G. Papandreou, I. Kokkinos, K. Murphy and A. L. Yuille. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. 2015. *arXiv preprint arXiv:1606.00915*.
- [7] Tyle ek, R. and R. Šára. Spatial pattern templates for recognition of objects with regular structure. Pages 364–374 in *German Conference on Pattern Recognition*, 2013.
- [8] Teboul, O., L. Simon, P. Koutsourakis and N. Paragios. Segmentation of building facades using procedural shape priors. Pages 3105–3112 in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010.
- [9] Kor, F. and W. Förstner. eTRIMS image database for interpreting images of man-made scenes. Technical report TR-IGG-P-2009-01, University of Bonn, Dept. of Photogrammetry, 2009.
- [10] Gadde, R., V. Jampani, R. Marlet and P. V. Gehler. Efficient 2D and 3D facade segmentation using auto-context. 2017. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- [11] Martinović, A., J. Knopp, H. Riemenschneider and L. Van Gool. 3D all the way: Semantic segmentation of urban scenes from start to end in 3D. Pages 4456–4465 in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015.
- [12] Lin, Y., M. Y. Yang and F. Nex. Semantic building façade segmentation from airborne oblique images. In *ISPRS TC II Mid-term Symposium*, 2018.
- [13] Martinović, A. and L. Van Gool. Bayesian grammar learning for inverse procedural modeling. Pages 201–208 in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, 2013.
- [14] Li, W. and M. Y. Yang. Efficient semantic segmentation of man-made scenes using fully-connected conditional random field. 2016. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences* 41.
- [15] Rahmani, K., H. Huang and H. Mayer. Facade segmentation with a structured random forest. 2017. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences* 4.
- [16] Yang, M. Y. and W. Förstner. Regionwise classification of building facade images. Pages 209–220 in *Photogrammetric Image Analysis: ISPRS Conference, PIA 2011*, held in Munich, Germany, 5–7 October 2011. Edited by U. Stilla, F. Rottensteiner, H. Mayer, B. Jutzi and M. Butenuth vol. 6952 LNCS. Berlin, Heidelberg: Springer.
- [17] Krähenbühl, P. and V. Koltun. Efficient inference in fully connected CRFs with Gaussian edge potentials. 2011. *Advances in Neural Information Processing Systems*: 109–117.
- [18] Martinović, A., M. Mathias, J. Weissenberg and L. Van Gool. A three-layered approach to facade parsing. Pages 416–429 in *Computer Vision–ECCV 2012*, 2012.
- [19] Gevaert, C. M., C. Persello, R. Sliuzas and G. Vosselman. Informal settlement classification using point-cloud and image-based features from UAV data. 2017. *ISPRS Journal of Photogrammetry and Remote Sensing* 125: 225–236.
- [20] Vetrivel, A., M. Gerke, N. Kerle, F. Nex and G. Vosselman. Disaster damage detection through synergistic use of deep learning and 3D point cloud features derived from very high resolution oblique aerial images, and multiple-kernel-learning. 2017. *ISPRS Journal of Photogrammetry and Remote Sensing* (2017).
- [21] Fooladgar, F. and S. Kasaei. Semantic segmentation of RGB-D images using 3D and local neighbouring features. Pages 1–7 in *2015 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, 2015.
- [22] Yang, Z. et al. A convolutional neural network-based 3D semantic labeling method for ALS point clouds. 2017. *Remote Sensing* 9 (9):936.
- [23] Yang, Z., B. Tan, H. Pei and W. Jiang. Segmentation and multi-scale convolutional neural network-based classification of airborne laser scanner data. 2018. *Sensors* 18: (10):3347.
- [24] Liu, C., J. Yuen and A. Torralba. SIFT flow: Dense correspondence across scenes and its applications. 2011. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33 (5):978–994.
- [25] Lowe, D. G. Distinctive image features from scale-invariant keypoints. 2004. *International Journal of Computer Vision* 60 (2):91–110.
- [26] Varma, M. and A. Zisserman. A statistical approach to texture classification from single images. 2005. *International Journal of Computer Vision* 62 (1/2): 61–81.
- [27] Vosselman, G., M. Coenen and F. Rottensteiner. Contextual segment-based classification of airborne laser scanner data. 2017. *ISPRS Journal of Photogrammetry and Remote Sensing* 128: 354–371.
- [28] Brodu, N. and D. Lague. 3D terrestrial lidar data classification of complex natural scenes using a multi-scale dimensionality criterion: Applications in geomorphology. 2012. *ISPRS Journal of Photogrammetry and Remote Sensing* 68:121–134.
- [29] Boykov, Y., O. Veksler and R. Zabih. Fast approximate energy minimization via graph cuts. 2001. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23 (11): 1222–1239.
- [30] Breiman, L. Random forests. 2001. *Machine Learning* 45 (1):5–32.
- [31] Krizhevsky, A., I. Sutskever and G. E. Hinton. ImageNet classification with deep convolutional neural networks. Pages 1097–1105 in *Proceeding NIPS'12 Proceedings of the 25th International Conference on Neural Information Processing Systems*, 2012, vol. 1.
- [32] Everingham, M. et al. The PASCAL visual object classes (VOC) challenge. 2010. *International Journal of Computer Vision* 88 (2):303–338.
- [33] Liu, H., J. Zhang, J. Zhu and S. C. H. Hoi. DeepFacade: A deep learning approach to facade parsing. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, vol. IJCAI-17, 2017.

SUSTAININGMEMBERS

ACI USA Inc.

Weston, Florida
acicorporation.com
Member Since: 1/2018

Aerial Services, Inc.

Cedar Falls, Iowa
www.AerialServicesInc.com
Member Since: 5/2001

Airbus Defense and Space

Chantilly, Virginia
www.intelligence-airbusds.com
Member Since: 6/2016

Applanix

a Trimble Company
Richmond Hill, Ontario, Canada
www.applanix.com
Member Since: 7/1997

Axis GeoSpatial, LLC

Easton, Maryland
www.axisgeospatial.com
Member Since: 10/2002

Ayres Associates, Inc.

Madison, Wisconsin
www.AyresAssociates.com
Member Since: 1/1953

Bohannon Huston, Inc.

Albuquerque, New Mexico
www.bhinc.com
Member Since: 11/1992

Cardinal Systems, LLC

Flagler Beach, Florida
www.cardinalsystems.net
Member Since: 1/2001

Certainty 3D LLC

Orlando, Florida
www.certainty3d.com
Member Since: 3/2011

CompassData, Inc.

Centennial, Colorado
www.compassdatainc.com
Member Since: 1/2012

DAT/EM Systems International

Anchorage, Alaska
www.datem.com
Member Since: 1/1974

Digital Mapping, Inc. (DMI)

Huntington Beach, California
www.admap.com
Member Since: 3/2002

Dewberry

Fairfax, Virginia
www.dewberry.com
Member Since: 1/1985

DigitalGlobe, Inc.

Longmont, Colorado
www.digitalglobe.com
Member Since: 6/1996

Environmental Research Incorporated

Linden, Virginia
www.eri.us.com
Member Since: 8/2008

Esri

Redlands, California
www.esri.com
Member Since: 1/1987

GeoBC

Victoria, Canada
www.geobc.gov.bc.ca
Member Since: 12/2008

GeoCue Group

Madison, Alabama
info@geocue.com
Member Since: 10/2003

Geomni, Inc.

Lehi, Utah
Geomni.net/psm
Member Since: 03/2018

GeoWing Mapping Inc.

Oakland, California
www.geowingmapping.com
Member Since: 1/2017

Global Science & Technology, Inc.

Greenbelt, Maryland
www.gst.com
Member Since: 10/2010

GPI Geospatial Inc. formerly Aerial Cartographics of America, Inc. (ACA)

Orlando, Florida
www.aca-net.com
Member Since: 10/1994

Harris Corporation

Broomfield, Colorado
www.harris.com
Member Since: 6/2008

Keystone Aerial Surveys, Inc.

Philadelphia, Pennsylvania
www.keystoneaerialsurveys.com
Member Since: 1/1985

Kucera International

Willoughby, Ohio
www.kucerainternational.com
Member Since: 1/1992

Lead'Air, Inc.

Kissimmee, Florida
www.trackair.com
Member Since: 5/2001

LizardTech

Portland, Oregon
www.lizardtech.com
Member Since: 9/1997

Martinez Geospatial, Inc. (MTZ)

Eagan, Minnesota
www.mtzgeo.com
Member Since: 1/1979

Maser Consulting P.A.

Tampa, Florida
www.maserconsulting.com
Member Since: 4/2015

MDA Information Systems LLC

Gaithersburg, Maryland
www.mdaus.com
Member Since: 1/1983

Merrick & Company

Greenwood Village, Colorado
www.merrick.com/gis
Member Since: 4/1995

Microsoft UltraCam Team (Vexcel Imaging, GmbH)

Graz, Austria
www.microsoft.com/ultracam
Member Since: 6/2001

Michael Baker Jr., Inc.

Beaver, Pennsylvania
www.mbakercorp.com
Member Since: 1/1950

Observera, Inc.

Chantilly, Virginia
www.observera.com
Member Since: 8/2002

Office of Surface Mining

Denver, Colorado
www.tips.osmre.gov
Member Since: 8/2017

PCI Geomatics

Richmond Hill, Ontario, Canada
www.pcigeomatics.com
Member Since: 1/1989

Pickett and Associates, Inc.

Bartow, Florida
www.pickettusa.com
Member Since: 4/2007

PixElement

Columbus, Ohio
www.pixelement.com
Member Since: 1/2017

Quantum Spatial, Inc.

Sheboygan Falls, Wisconsin
www.quantumspatial.com
Member Since 1/1974

Riegl USA, Inc.

Orlando, Florida
www.rieglusa.com
Member Since: 11/2004

Robinson Aerial Survey, Inc. (RAS)

Hackettstown, New Jersey
www.robinsonaerial.com
Member Since: 1/1954

Routescene, Inc.

Durango, Colorado
www.routescene.com/
Member Since: 12/2007

Sanborn Map Company

Colorado Springs, Colorado
www.sanborn.com
Member Since: 10/1984

Sidwell Company

St. Charles, Illinois
www.sidwellco.com
Member Since: 11/1992

SkyIMD— Imaging Mapping & Data

Richmond, California
www.skyimd.com
Member Since: 1/2017

Surveying And Mapping, LLC (SAM)

Austin, Texas
www.sam.biz
Member Since: 12/2005

Teledyne Optech

Toronto, Canada
www.teledyneoptech.com
Member Since: 1/1999

Terra Remote Sensing (USA) Inc.

Bellevue, Washington
www.terraremsensing.com
Member Since: 10/2016

The Airborne Sensing Corporation

Toronto, Canada
www.airnsensing.com
Member Since: 7/2003

Towill, Inc.

San Francisco, California
www.towill.com
Member Since: 1/1952

University of Twente/Faculty ITC

Enschede, Netherlands
www.itc.nl
Member Since: 1/1992

U.S. Geological Survey

Reston, Virginia
www.usgs.gov
Member Since: 4/2002

Woolpert LLP

Dayton, Ohio
www.woolpert.com
Member Since: 1/1985

RoofN3D: A Database for 3D Building Reconstruction with Deep Learning

Andreas Wichmann, Amgad Agoub, Valentina Schmidt, and Martin Kada

Abstract

Machine learning methods, in particular those based on deep learning, have gained in importance through the latest development of artificial intelligence and computer hardware. However, the direct application of deep learning methods to improve the results of 3D building reconstruction is often not possible due, for example, to the lack of suitable training data. To address this issue, we present RoofN3D which provides a three-dimensional (3D) point cloud training dataset that can be used to train machine learning models for different tasks in the context of 3D building reconstruction. The details about RoofN3D and the developed framework to automatically derive such training data are described in this paper. Furthermore, we provide an overview of other available 3D point cloud training data and approaches from current literature in which solutions for the application of deep learning to 3D point cloud data are presented. Finally, we exemplarily demonstrate how the provided data can be used to classify building roofs with the PointNet framework.

Introduction

In recent years, machine learning has been extensively studied with the aim of automatically generating models from data. For this purpose, several machine learning methods have been developed that do not simply memorize examples but are able to automatically recognize patterns and rules in the training data. Particularly approaches based on deep learning have achieved excellent results for different applications. For image classification tasks, for example, deep learning methods using convolutional neural network (CNN) architectures have become a standard framework in the last few years, as their results are already comparable or even better than from human experts (Ciresan *et al.* 2012; Krizhevsky *et al.* 2012).

While CNNs have been a great success for images, they have been, however, less successful for three-dimensional (3D) point clouds. The reasons for this are manifold, but they can be mainly attributed either to the lack of publicly available training data or to the specific properties of point clouds: (i) Point cloud data is unstructured and not gridded, which is why convolutional kernels cannot be directly applied; (ii) Point cloud data usually has a varying point density, which can lead to a weak generalization of the learned feature representations; (iii) Besides intensity, there is often no radiometric information (e.g., color) available.

Although many different solutions have been already proposed in recent years to apply deep learning to 3D point clouds, a major drawback that still remains is the lack of publicly available training data that can be used to train neural networks. Since machine learning techniques usually require a large amount of training data, this issue is very crucial for carrying out research in this area. While there is a large amount of training data for images available (e.g., ImageNet

(Deng *et al.* 2009), Modified National Institute of Standards and Technology (MNIST) (Deng 2012), CIFAR10/CIFAR100 (Krizhevsky 2009)), the amount of training data for 3D point clouds is comparatively small.

By evaluating a large number of publicly available 3D point cloud training datasets for machine learning, it became clear that a good basis for traditional classification tasks is already available. In comparison to available image databases, however, the number of classes is generally still quite limited. Particularly for buildings, which play an essential role for most applications in urban areas, we discovered a shortage in the number of available datasets. According to our knowledge, there is currently no publicly available 3D point cloud training dataset that provides distinct classes for buildings. However, many applications such as facility management and preservation of townscapes require a fine subdivision of the building class, for example, to distinguish between different roof types or to recognize certain roof structures. In order to close this crucial gap, we have developed an automatic workflow in which building points of an aerial LiDAR dataset are processed so that they can be used to train deep neural networks for different tasks in the context of 3D building reconstruction. For this purpose, only data for those buildings are provided for which the workflow achieves good results. At this point, the data of all other buildings are discarded with the intention that they can later be better processed through a reconstruction approach based on deep learning.

The proposed workflow has been applied to the publicly available New York City (NYC) dataset which consists of over one million buildings. The resulting training dataset is made available through RoofN3D and provides not only geometric information but semantic information as well. Note, RoofN3D provides only training data and is not a benchmark dataset at the moment. The training data is publicly available at <https://roofn3d.gis.tu-berlin.de>.

The remainder of the paper is organized as follows: First, recent deep learning methods from the literature are summarized in which solutions are proposed to overcome the aforementioned issues related to the specific properties of point clouds. In this context, general trends for current and future research directions are pointed out. Afterward, an overview of publicly available 3D point cloud training data for machine learning is presented. Subsequently, details of the developed workflow and the training data provided by RoofN3D are described. To demonstrate the suitability of the provided training data for deep learning algorithms, we exemplarily show how the data can be used to train the PointNet framework with the purpose of classifying building roofs. Finally, a conclusion and potential future enhancements concerning RoofN3D are presented.

Photogrammetric Engineering & Remote Sensing
Vol. 85, No. 6, June 2019, pp. 435–443.
0099-1112/19/435–443

© 2019 American Society for Photogrammetry
and Remote Sensing
doi: 10.14358/PERS.85.6.435

Institute of Geodesy and Geoinformation Science (IGG),
Technische Universität Berlin, Straße des 17. Juni 135, 10623
Berlin, Germany (andreas.wichmann@tu-berlin.de).

Deep Learning for 3D Point Clouds

The automatic recognition of objects is a fundamental task in computer vision. It has recently attracted again the interest of many researchers due to the advancements in artificial intelligence and the development of hardware which enables the implementation and the application of deep neural networks. The results that can be achieved with deep neural networks have already reached a new level for two-dimensional (2D) images. Particularly CNNs have proven to be capable of providing state-of-the-art results; see, for example, AlexNet (Krizhevsky *et al.* 2012), VGG (Simonyan and Zisserman 2015), GoogLeNet (Szegedy *et al.* 2015), and ResNet (He *et al.* 2016). Therefore, current research is carried out to apply CNNs not only on 2D data but also on 3D data such as point clouds. The essence of a CNN is, however, the convolutional layer whose parameters are represented by a set of kernels (also called filters). These convolutional kernels enable, on the one hand, to share weights in convolutional layers and thus significantly reduce the total number of parameters in a CNN. On the other hand, convolutional kernels always require data in a regular structure as input. Therefore, CNNs cannot be directly applied to 3D point cloud data which consist of a set of unordered points. For further information about CNNs, see, for example, Goodfellow *et al.* (2016).

To overcome this issue and to adequately deal with the sparsity problem of three-dimensional point cloud data, solutions have been developed that are based on the conversion of the irregular point cloud data structure to a regular data structure. Some proposed regular representations in the context of CNNs with respect to 3D point cloud data are described in the section "Regular Representations". In addition to approaches based on regular representations, further approaches for neural networks have been developed that are able to directly process data represented in an irregular data structure. Some details of these irregular representations are summarized in the section "Irregular Representations".

Regular Representations

A solution that has been proposed to represent 3D point cloud data in a regular structure, so that they are suitable for CNNs, is the multi-view representation. In this approach, multiple 2D views (i.e., images) of a 3D object are generated from different viewpoints. Thereby, the dimension of the 3D input data is reduced for each viewpoint to 2D. An important aspect in this context is how the viewpoints can be determined. Different approaches have been developed for multi-view CNNs in which the number and the locations of the camera are fixed empirically (Su *et al.* 2015) or in which the optimal camera locations are learned (Kalogerakis *et al.* 2017; Huang *et al.* 2018). Once a set of 2D images has been captured from different viewpoints, each image is processed by a CNN and the resulting feature maps are aggregated by applying view pooling. Further examples of different multi-view CNNs are presented in Qi *et al.* (2016) and Yi *et al.* (2017).

While multi-view representations provide the advantage of being compatible with 2D CNNs, shape details may be obscured during the conversion process. Particularly occlusions and a limited number of viewpoints can exacerbate this problem. All this might result in inconsistencies during the reconstruction of surfaces in 3D space.

Another group of regular representations, which keep the dimensions of the input data, is the volumetric representation. In the last years, different variants of volumetric representations have been proposed for CNNs. A well-known representative is the voxel grid. In a voxel grid, the three-dimensional space is discretized into a regular grid and each resulting voxel is assigned a value based on the points within the voxel. A special type of a voxel grid, which is commonly in use, is the occupancy grid (Thrun 2001). The special

characteristic of an occupancy grid is that each voxel is assigned only the value occupied or unoccupied, depending on the presence or absence of data. Some examples in which different variants of voxel grids have been used as input for CNNs are VoxNet (Maturana and Scherer 2015), 3D ShapeNets (Wu *et al.* 2015), volumetric CNNs (Qi *et al.* 2016), SEG-Cloud (Tchapmi *et al.* 2017), and VoxelNet (Zhou and Tuzel 2018). In addition to the occupancy values, voxel grid cells may store further information such as statistics of the points within a cell (Song and Xiao 2014), distances from the cell center to the nearest point (Song and Xiao 2016), or features extracted from the local neighborhoods (Wang and Posner 2015; Engelcke *et al.* 2017).

A major advantage of representing 3D point clouds in a voxel grid is that already existing CNN architectures for images can be generally applied with only a few adaptations. However, since objects are represented in 3D point cloud data only on their surface, the input data easily become unnecessarily voluminous due to a large number of unoccupied voxels; see, for example, Li *et al.* (2016). An illustration of this so-called sparsity problem of 3D data in occupancy grids is shown in Figure 1 for a hip roof in different resolutions. In order to overcome this problem, while keeping the spatial information about the 3D shape, other regular data structures have been developed such as deep data structures and convolutional filters that can work on them.

In deep data structures, the memory usage is significantly reduced by recursively dividing only occupied voxels. An illustration of a deep data structure for a saddleback roof in different resolutions is shown in Figure 2. Thus, deep data structures enable in practice the use of a higher grid resolution compared to occupancy grids. Some examples in which deep data structures have been used are given in Riegler *et al.* (2017), Klovov and Lempitsky (2017), and Wang and Posner (2015).

A general drawback of regular data structures is that the resolution to represent 3D data is generally limited. Therefore, the conversion of irregular to regular 3D data structures is usually accompanied by information loss. Furthermore, for adequately capturing some shape details such as symmetry and roundness, very high-resolution 3D grids are required, which significantly increase the memory consumption and the computational cost. Designing machine learning architectures that can directly process irregular 3D data can help to overcome these challenges and also the data sparsity problem.

Irregular Representations

In addition to regular representations, other solutions have been proposed for deep learning with respect to 3D point clouds based on irregular representations. Early approaches for irregular representations have been presented in Fang *et al.* (2015) and Guo *et al.* (2015). In the former, shape descriptors are used which enable the identification of a 3D object as a member of a category based on a concise and informative representation. Thereby, 3D input meshes are represented in 2D without losing relevant information. In the latter, a compact mesh representation is learned by extracting multiple geometric low-level features. Based on the extracted features, a 2D tensor is constructed that serves as input for a CNN model.

An established neural network that accepts point cloud data in their original format is PointNet (Qi *et al.* 2017a). The distinct feature of PointNet is its capability to comply with the permutation invariance, a fundamental property of point cloud data, by employing symmetric functions. Thus, points are individually processed by a sequence of multilayer perceptron layers (MLP layers). The resulting higher dimensional representations are aggregated to a global representation by exploiting the symmetry property of the max pooling function, which is applied across each dimension of the feature space. Additionally, alignment networks are integrated into

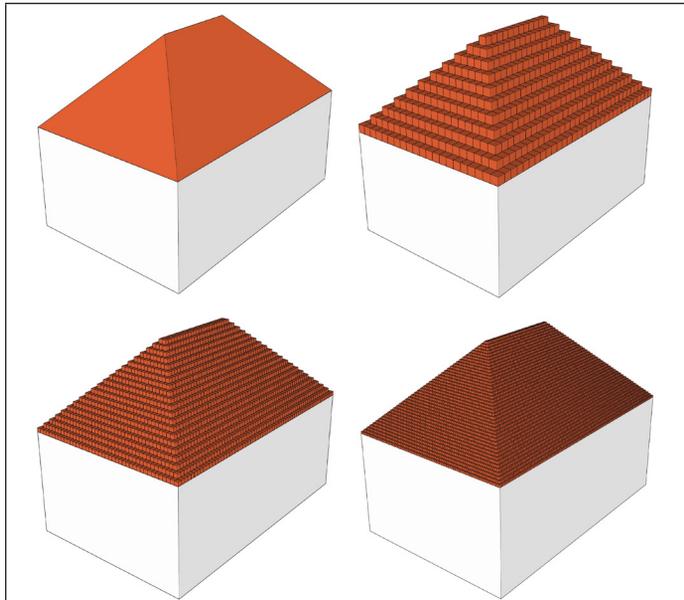


Figure 1. The percentage of occupied cells of the shown hip roof in a volumetric representation at resolution 32^3 is 2.506% and reduced to 1.025% and 0.513% at resolution 64^3 and 128^3 , respectively.

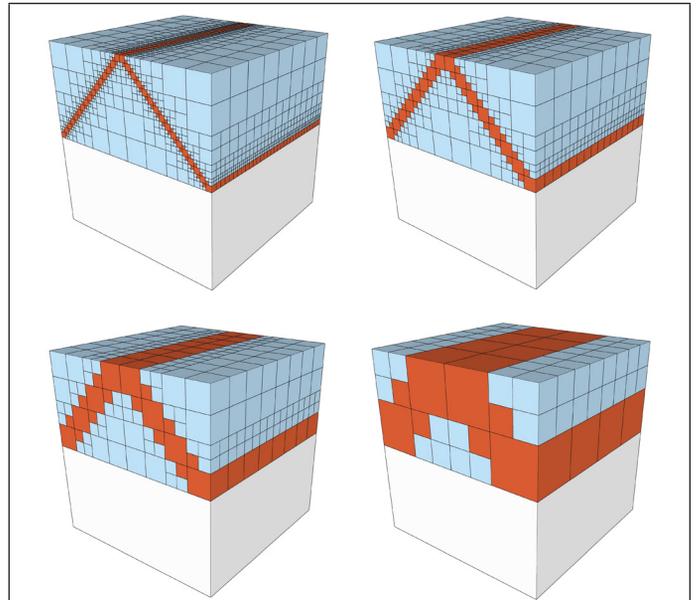


Figure 2. A saddleback roof represented in a deep data structure at resolution 64^3 , 32^3 , 16^3 , and 8^3 . Occupied cells are colored in red, while unoccupied are colored in blue. Unoccupied cells above and below the hip roof were omitted for reasons of clarity.

the network architecture to boost the performance of the classifier. An alignment module (spatial transformer) regresses the transformation parameters of the original object pose (intermediate feature maps) with respect to a canonical pose, and subsequently applies the learned transformation to correct the original or intermediate orientation.

Compared to convolutional models that benefit from hierarchical feature learning, PointNet has no spatial context knowledge, besides the individually extracted point representations and the global point cloud representation. However, the positive effect of leveraging spatial correlations when processing geodata have already been shown in some approaches such as Gressin *et al.* (2013) and Weinmann *et al.* (2015). PointNet++ (Qi *et al.* 2017b), an upgraded version of PointNet, addresses the spatial context knowledge by integrating sampling and grouping layers into the network. The neighborhoods defined by these layers are processed by applying the standard version of PointNet.

A more recent approach to irregular representations implements a convolutional operator in which the nearest point neighbors are queried on the fly in a pointwise convolution (Hua *et al.* 2018). In this context, a preprocessing step has been proposed in Corcoran *et al.* (2018), in which the point cloud is projected to a space where the point locations are in a certain order. By projecting and ordering the points, for example, on a 3D space-filling one-dimensional curve such as a Hilbert curve, the direct use of one- or two-dimensional convolutions on the point cloud data is enabled.

Point Cloud data can be used not only with discriminative models but also with generative ones, as for instance with Generative Adversarial Network (GAN) (Goodfellow 2016). They enable the generation of new data from a given distribution using an adversarial training paradigm between a discriminator and a generator network. As an example, the generator model in 3D-GAN (Wu *et al.* 2016) can generate 3D models from a given statistical distribution while the discriminator model is used as 3D descriptor for object recognition tasks. In Achlioptas *et al.* (2018), an Auto Encoder (AE) network is designed to be invariant to the permutations and

spatial transformations of its input following the concepts proposed by the PointNet architecture. The higher dimensional data representation learned by the AE is then used for an improved GAN training compared to only when using original data.

Available 3D Point Cloud Training Data

In machine learning, training data is usually needed for learning a model such as a neural network. A challenge for the training of complex models is the provision of a sufficient amount of training data. If a complex model is trained on an insufficient amount of training examples, the model will have difficulties to generalize. Thus, the availability of massive training data is nowadays of high importance. Another important aspect of the availability of training data is that it can be also used to evaluate machine learning models. Many machine learning models are constructed as very complex mathematical models, which can make a theoretical evaluation of their performance very challenging and controversial. However, a practical approach to address this challenge is to empirically evaluate machine learning models within a common framework based on publicly available data. Empirical evaluations provide researchers the ability to compare and evaluate the performance of their models according to set standards. Early publicly available datasets have been proven to be of great value to research. A frequently used dataset is MNIST (Deng 2012), which can be used to prove the effectiveness of a model; see, for example, (Sabour *et al.* 2017). Generally, it is noticeable that many research communities are nowadays interested in acquiring and producing such data in 3D. With ShapeNet (Chang *et al.* 2015) and ModelNet (Wu *et al.* 2015), for example, datasets of 3D models from common 3D objects are publicly available. Some further publicly available 3D datasets are listed in the following paragraph.

There are, nowadays, many 3D datasets available that were captured in urban areas with the help of mobile laser scanning. This includes, for example, data from street sections such as the Paris-rue-Madame dataset (Serna *et al.* 2014),

which consists of about 10 million points and 27 classes, or the Paris-rue-Soufflot dataset (Hernández and Marcote-gui 2009), which consists of about 20 million points and 6 classes. But also data of larger urban areas are available, such as the IQumulus & TerraMobilita Contest dataset (Vallet *et al.* 2015), which consists of about 300 million points and about 80 classes, and the Semantic3D.Net dataset (Hackel *et al.* 2017), which consists of over four billion manually labelled points and eight classes.

The provided datasets listed in the previous paragraph are generally suitable for traditional tasks such as recognition and classification in the urban context. However, for the purpose of 3D building type classification and reconstruction, these datasets would be not sufficient because they do not provide distinct subclasses within the building class which indicate the building parts. The Oakland dataset (Munoz *et al.* 2009), which consists of about 1.6 million points and 44 classes, can also only partially overcome this issue. It contains classes for building parts and a specific label for roofs, but it does not provide an explicit class distinction between different roof types.

To the best of our knowledge, there is currently no publicly available dataset with semantic roof types for the purpose of learning different roof types and 3D building reconstruction based on machine learning models. By presenting an automatic workflow that can provide such training datasets and by providing the results of already processed data on RoofN3D, we aim to close this crucial gap and we would be pleased about the use of this data in other research work.

Training Data Preparation Framework

For the provision of massive 3D data and further information that are needed to train deep neural networks in the context of 3D building reconstruction, an automatic framework has been developed. The automatic framework has been designed to be capable of efficiently processing very large point clouds with billions of points. Thereby, huge training datasets consisting of different building classes and a large number of instances for each class can be generated in fairly short time. An overview of the whole framework is illustrated in Figure 3. It consists of the following two major steps: (i) extraction of building points and (ii) derivation of building information.

In the building point extraction step, aerial point clouds and building footprints are used as input data for the developed framework. Both types of data are nowadays already

publicly available for several cities, states, and countries around the world (e.g., New York City (USA), Philadelphia (USA), Toronto (Canada), Vancouver (Canada), Cambridge (UK), North Rhine-Westphalia (Germany), Thuringia (Germany), The Netherlands, etc.). Due to the large number of points in the provided point clouds, a direct extraction of building points is usually not feasible in a reasonable time if, for example, only building footprints in combination with a ray casting algorithm are used to solve the point-in-polygon problem. Therefore, point clouds and building footprints are not directly intersected with each other to determine the building points but multiple patches are first generated for each point cloud. The resulting patches are then intersected with those building footprints that are located in the area of the point cloud. In this way, all patches are identified that feature an overlap with a building. Afterward, the relevant patches are exploded and all those points are classified as building points that are located within a building footprint of the processed area. The classified points and the building footprints are finally stored in the RoofN3D database.

The derivation of building information based on building points follows the principles of common data-driven building reconstruction approaches and consists of the following three sub-steps: (i) segmentation of roof surfaces, (ii) shape recognition, and (iii) 3D building model construction. Since the results of each sub-step can be used to train neural networks for different tasks, they are stored in the RoofN3D database.

For the segmentation of planar roof surfaces on the basis of building points, subsurface growing is applied as described in Kada and Wichmann (2012). In contrast to the well-known surface growing method, segments are enabled during the subsurface growing procedure to grow below the surface. As a result, segment patches that belong to the same roof surface, but are disconnected by roof superstructures, are merged together. Consequently, symmetries are implicitly modeled and the number of primitives is reduced of which a building with complex roof structure is comprised of. Furthermore, segments of connected building components (e.g., dormer and base roof) feature actual intersections so that gaps between them are automatically closed. The advantages of subsurface growing support the identification of roof structures in the subsequent shape recognition step and make their detection more robust than with conventional surface growing.

In order to further improve the segmentation results, subsurface growing has been extended by a reassignment

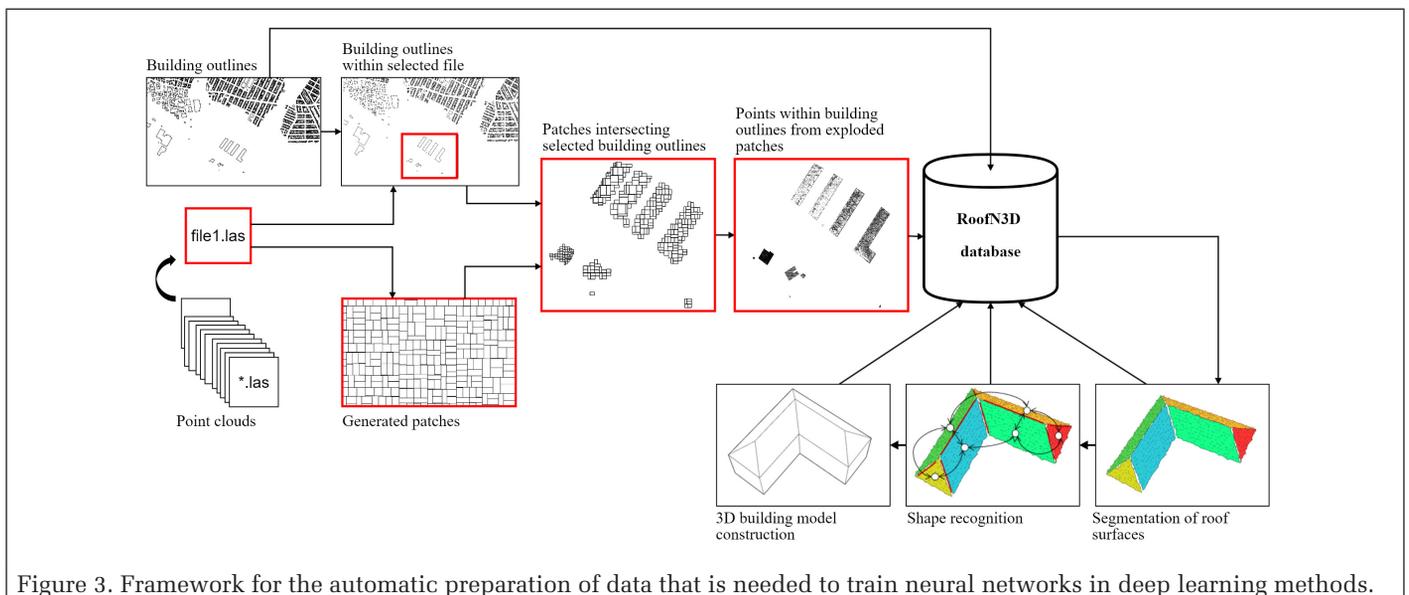


Figure 3. Framework for the automatic preparation of data that is needed to train neural networks in deep learning methods.

method as suggested in Vosselman and Klein (2010). In this method, all points assigned to a segment are reassigned to a neighboring segment in cases where they would better fit the neighboring segment in terms of distance and normal vector direction. The reassignment process is carried out once after surface growing and once after subsurface growing. In addition to the improved assignment of points that already belong to a segment, subsurface growing has also been extended in such a way that already segmented roof surfaces are further enriched with points not previously assigned to any segment. For this, a point in the set of unassigned points is assigned to its closest segment after the reassignment process has been completed if its distance to the segment is within a certain tolerance. Both described extensions are suitable to improve the segmentation result of subsurface growing.

Once planar roof surfaces have been determined, all segments are represented as nodes and their adjacency relationships as connecting edges in a so-called roof topology graph (RTG). For the recognition of certain roof shapes in the RTG, a graph grammar has been developed in which production rules are defined that represent possible graph transformations. Thereby, the search for predefined roof shape models does not need to be performed directly on the input data but it can be carried out on higher-level information in the so-called topology space. Thus, the robustness of traditional model-driven recognition approaches is maintained while reducing the search effort and the computational time.

Each production rule of the developed grammar consists of two graphs representing its left-hand side (LHS) and its right-hand side (RHS). If a production rule is applied to the RTG, all occurrences of the LHS in the RTG are first identified by a labeled graph matching algorithm and then replaced by its RHS. The production rules have been essentially designed in such a way that adjacent nodes and their connecting edges, which represent lower-level information, are aggregated to a single node which represents higher-level information about the building. For example, two connected nodes that both represent sloped segments, whose segment normals point in the horizontal plane in the opposite direction, and which have an intersection line of a certain minimum length are aggregated to a single node that represents the semantic information of a saddleback roof. With each aggregation, semantic information is added to the RTG. By applying several production rules of the graph grammar to the RTG, higher-level information about the building to be reconstructed is derived. Thereby, the knowledge of the building structure including the building parts becomes available. To ensure that unnatural shapes are automatically discarded, already derived building knowledge is incorporated during the application of production rules. Due to the expressive power of the applied grammar, not only geometric information but also extensive semantic information can be provided.

Finally, 3D building models are constructed based on half-space modeling as introduced in Kada and Wichmann (2013) and adjusted based on the divisive clustering techniques introduced in Wichmann and Kada (2014) to support natural building characteristics (e.g., symmetry, coplanarity, orthogonality). By utilizing half-space modeling, buildings are represented within the proposed framework as closed solids to guarantee that all building models are topologically correct and that they do not feature any unwanted gaps or intersections.

RoofN3D

In order to close the gap of publicly available 3D point cloud data that are suitable to train neural networks for different tasks in the context of buildings, we present RoofN3D. It provides a platform for the distribution of 3D training data

that result from the application of the presented training data preparation framework to various data. As a first step, the publicly available New York City dataset of the NYC Open Data Portal (<https://opendata.cityofnewyork.us>) has been processed and suitable parts of the resulting training data are made available via RoofN3D. Some information about the New York dataset are summarized in the section “New York Dataset” and further details about the provided data on RoofN3D are given in the section “RoofN3D Data”. Note, further datasets will be processed in the future and their results will be added to RoofN3D. The training data is available at <https://roofn3d.gis.tu-berlin.de>.

New York Dataset

The building footprint dataset of New York is part of the planimetrics geodatabase and used by the NYC Department of Information Technology and Telecommunications Geographic Information Systems group to maintain and distribute an accurate base map for NYC. They are derived from images of the 2014 New York Statewide Flyover, which includes raw imagery collected to support the generation of 0.5 feet ground sampling distance natural color imagery. The provided building footprints represent the perimeter outline of each building. Divisions between adjoining buildings are determined by tax lot divisions. The estimated positional accuracy for 95% of the data is ± 2 feet. The whole dataset consists of more than one million building footprints.

The LiDAR point clouds of New York are provided by the U.S. Geological Survey. They have been captured from August 2013 to April 2014 and cover an area of 1,009.66 km². The average density of the point clouds is about 4.72 points/m².

RoofN3D Data

The available data from RoofN3D currently consist of the results of the presented training data preparation framework that has been applied to the New York dataset. The New York dataset has been selected because it provides a large number of buildings with a wide intraclass variation, which enables the training of complex models without being in risk of overfitting. An overview of the underlying architecture of RoofN3D and the available information about the buildings are shown in Figure 4. As illustrated, for each building in RoofN3D, a building footprint, all building points therein, segmentation results, semantic information, and geometric information are provided. Please note, the structure of RoofN3D is not fixed and might be adapted according to future needs.

The provided data can be used in various ways. For example, data resulting from the segmentation process can be used to train semantic segmentation networks. For this, each extracted segment provides the information about its assigned points, the outline, and the plane equation which best fits the assigned points. We have limited the segmentation process to planar areas because most roof surfaces can be described with sufficient accuracy by planar faces. This is for most buildings, particularly in residential areas, the case and useful because arbitrary shapes usually require computationally intensive surface fitting procedures (Wang 2013). Since subsurface growing has been implemented as an extension of the well-known surface growing method, segments of both surface growing and sub-surface growing are provided. For the latter, a distinction is made between surface points and virtual subsurface points which have been not initially captured but introduced to close unwanted gaps in a roof surface. In addition to the points that have been assigned to a segment, also all unassigned building points are provided which do not belong to any segment according to the applied surface growing and subsurface growing method. The outline of a segment has been determined by projecting all segment points onto

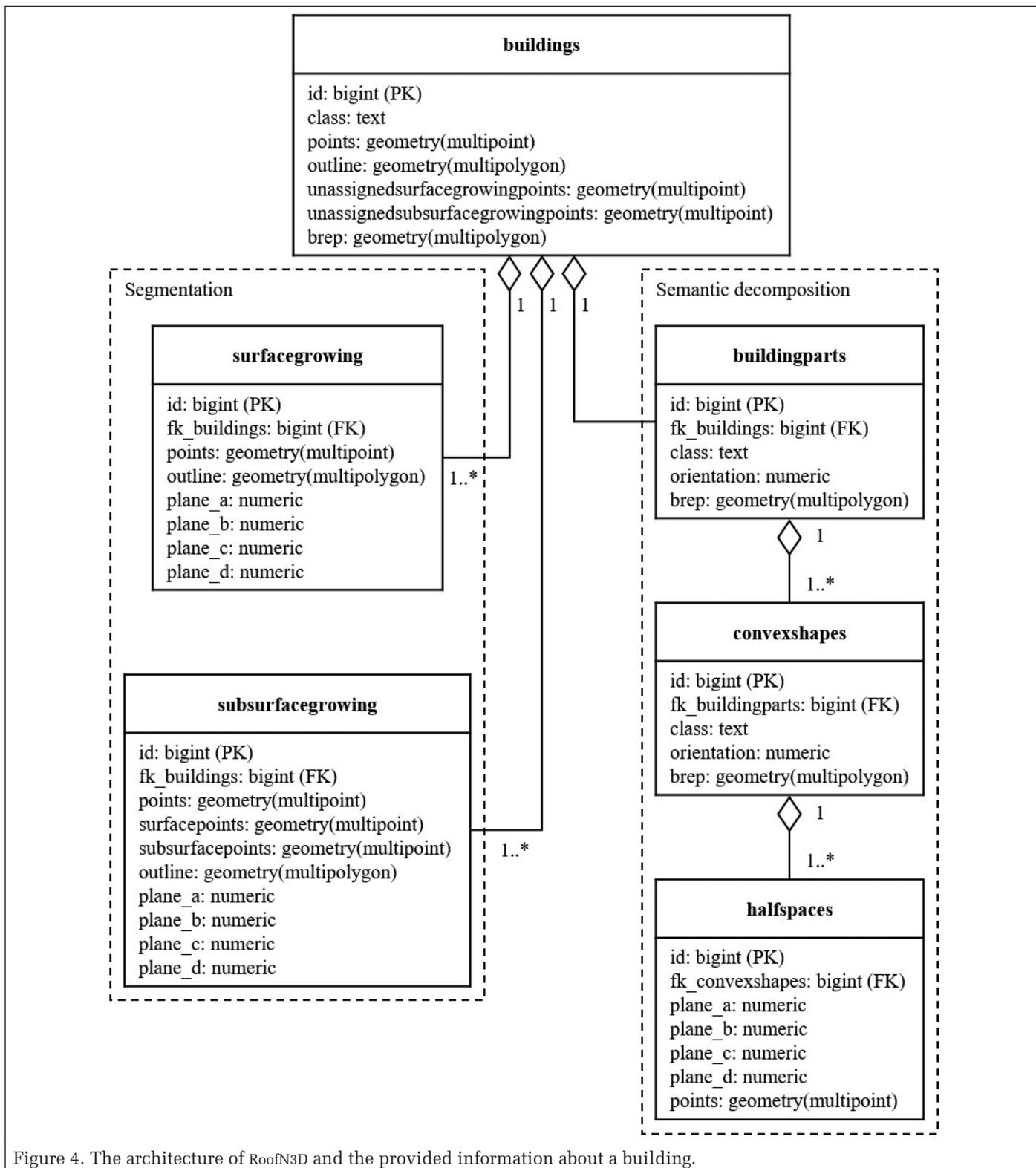


Figure 4. The architecture of RoofN3D and the provided information about a building.

the plane of the segment and by performing the alpha shape algorithm presented in Edelsbrunner *et al.* (1983).

To learn the structure of buildings with deep learning methods, the applied grammar of the shape recognition step has been designed in such a way that structural information about the buildings are derived. This includes the information about the building parts that compose the building. Depending on the complexity of the building, the number and the shape of the building parts can be very diverse. In order

to provide further information about those building parts that have a concave shape, information about convex components that compose the concave building part are provided. To reduce the processing effort for training a deep neural network, also information about the orientation of building parts and their convex components are provided.

Another important task that can be approached by means of neural networks and the provided training data is the learning of the geometric construction of 3D building models.

For this, RoofN3D provides boundary models not only for the whole building but also for its building parts. These boundary models are derived from the conversion of the closed solid that results from the applied automatic reconstruction method. Thereby, it can be guaranteed that the boundary models are always closed and represented as two-manifold polyhedrons. Note, due to the lack of ground level height information, all building models have been extruded to the same ground height (i.e., 0 m). In addition to boundary models, information about half-space models are provided via RoofN3D. For this, the plane equations of the hyperplanes that define the half-spaces of the roof of a convex building component are provided. By applying the Boolean intersection operator to these half-spaces, the roof shape of the convex component is defined. Furthermore, the roof shape of a building part can be derived by applying the Boolean union operator to the half-space representations of all convex components that compose the building part. Analogous, the roof shape of the whole building is given by uniting all half-space models of those building parts that compose the building. If the shape of the whole building with extruded facades is needed, the latter half-space model needs to be intersected once with a half-space having a horizontal hyperplane whose normal vector is pointing downwards and once with the provided building footprint formulated as a half-space model.

The aforementioned information is offered via RoofN3D for different types of roofs. First, the number of different roof types is limited and only cover simple shapes such as gable roofs, two-sided hip roofs, pyramid roofs, etc. However, more complex roof types will be added over time.

Experiments

To demonstrate the suitability of RoofN3D for deep learning algorithms, we exemplarily used the PointNet framework to classify building roofs based on the presegmented point cloud data of RoofN3D. We selected PointNet because it is one of the main neural network architectures that processes point cloud data in their original data structure. In the standard version of PointNet, the coordinates of each point are individually processed by a sequence of five MLP layers (output dimensions: 64, 64, 64, 128, 1024) implemented as 1 × 1 convolutions. Additionally, the orientations of the point clouds are corrected with regard to canonical orientations. For this, an alignment module is applied on the input coordinates in a first iteration and once more by using a second alignment module on the feature maps resulting from the first two MLP layers. The subsequent max pool layer summarizes the global representation of the point cloud. The global descriptor is further processed by a sequence of three fully connected layers before a classification layer is applied.

Although PointNet is generally able to deal with irregular data structures, some preprocessing steps were necessary to train the network with the RoofN3D data: Firstly, we limited the number of instances in each building class to the number of instances of the least represented class (i.e., pyramid roof). Secondly, we adjusted the number of points representing a building roof to a fixed value. Considering the low point density of about 5 points/m² and the correlation between the number of points and the size of the building, under-sampling would be at risk to discard so-called key points, which represent relevant information such as corners or edges (Qi *et al.* 2017a). Therefore, we selected training point clouds containing maximum 1024 points and over-sampled those with a lower number to the size of 1024 by randomly reusing existing points of the point cloud. Since the coordinates of each point in PointNet are individually projected into a higher dimensional space and the global signature of the point cloud

is calculated by applying max pooling across each feature, the artificially introduced redundancy has no effect on the global description of the roof shape and only slightly increases the computational effort.

For the training of the PointNet network, we normalized each point cloud by projecting them on a sphere of radius one and represented each point only by its 3D coordinate which we encoded in separate channels. We used in total 4620 examples, which were equally distributed over the classes saddle-back roof, two-sided hip roof, and pyramid roof, divided the dataset into a training set (3300 examples) and a test set (1320 examples), and fed them to the network in batches of 32 examples. The dataset augmentation was performed during training by applying a set of rotations around the vertical axis and by jittering the points with Gaussian noise (0, 0.2). The scores over the possible classes were computed by optimizing the cross-entropy loss. The model was trained using the Adam optimizer (Kingma and Ba 2015) with a learning rate of 0.001 for 250 epochs. The obtained average accuracy on the evaluation set was 94%.

In contrast to small objects (e.g., from the ShapeNet40 database) which have arbitrary orientations, building roofs deviate in most cases from canonical orientations only by a rotation around the vertical axis. Therefore, we further investigated the impact of the alignment modules by using the same training protocol as described above, but removing the alignment modules. The result was a 4% decrease in average accuracy.

Conclusion and Outlook

The training dataset available on RoofN3D provides aerial LiDAR data and building information that can be used to train deep neural networks for different tasks in the context of 3D building reconstruction. It has been exemplarily demonstrated how the provided presegmented point cloud data can be used to train the PointNet framework for the classification of building roofs. The training dataset has been recently published and will be extended in the future according to the needs of the research community. This includes, for example, the addition of further buildings with the same or other roof shapes. For the latter, buildings with complex roof shapes (e.g., featuring roof superstructures) are of particular interest. Additionally, it is important that the number of examples become more evenly distributed among the classes in the future. Furthermore, it is planned to carry out a quality assessment and to continuously improve the offered data. This is necessary because the training data was generated with an automatic process. We hope that the 3D training data on RoofN3D can be used in the future to further improve automatic 3D building reconstruction approaches and their results with various methods of machine learning.

Acknowledgment

We are grateful to the city of New York for providing building footprints through the New York City Open Data Portal (<https://opendata.cityofnewyork.us>) and to the U.S. Geological Survey (USGS) for providing LiDAR point clouds of New York.

References

- Achlioptas, P., O. Diamanti, I. Mitliagkas and L. Guibas. 2018. Learning representations and generative models for 3D point clouds. In *Proceedings of the 6th International Conference on Learning Representations (ICLR)*.
- Chang, A. X., T. Funkhouser, L. Guibas, P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, H. Su, J. Xiao, L. Yi and F. Yu. 2015. ShapeNet: An information-rich 3D model repository. *arXiv preprint arXiv:1512.03012*.

- Cirean, D., U. Meier and J. Schmidhuber. 2012. Multi-column deep neural networks for image classification. Pages 3642–3649 in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Corcoran, T., R. Zamora-Resendiz, X. Liu and S. Crivelli. 2018. A spatial mapping algorithm with applications in deep learning-based structure classification. *arXiv preprint arXiv:1802.02532v2*.
- Deng, J., W. Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei. 2009. ImageNet: A large-scale hierarchical image database. Pages 248–255 in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Deng, L. 2012. The MNIST database of handwritten digit images for machine learning research. *IEEE Signal Processing Magazine* 29(6):141–142.
- Edelsbrunner, H., D. Kirkpatrick and R. Seidel. 1983. On the shape of a set of points in the plane. *IEEE Transactions on Information Theory* 29(4): 551–559.
- Engelcke, M., D. Rao, D. Z. Wang, C. H. Tong and I. Posner. 2017. Vote3Deep: Fast object detection in 3D point clouds using efficient convolutional neural networks. Pages 1355–1361 in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*.
- Fang, Y., J. Xie, G. Dai, M. Wang, F. Zhu, T. Xu and E. Wong. 2015. 3D deep shape descriptor. Pages 2319–2328 in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Goodfellow, I. 2016. NIPS 2016 tutorial: Generative adversarial networks. *arXiv preprint arXiv:1701.00160v4*.
- Goodfellow, I., Y. Bengio and A. Courville. 2016. *Deep Learning*. Cambridge, Mass.: The MIT Press.
- Gressin, A., C. Mallet, J. Demantké and N. David. 2013. Towards 3D LiDAR point cloud registration improvement using optimal neighborhood knowledge. *ISPRS Journal of Photogrammetry and Remote Sensing* 79:240–251.
- Guo, K., D. Zou and X. Chen. 2015. 3D mesh labeling via deep convolutional neural networks. *ACM Transactions on Graphics (TOG)* 35(1):3.
- Hackel, T., N. Savinov, L. Ladicky, J. D. Wegner, K. Schindler and M. Pollefeys. 2017. Semantic3D.Net: A new large-scale point cloud classification benchmark. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences IV-1/W1:91–98*.
- He, K., X. Zhang, S. Ren and J. Sun. 2016. Deep residual learning for image recognition. Pages 770–778 in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Hernández, J. and B. Marcotegui. 2009. Filtering of artifacts and pavement segmentation from mobile LiDAR data. *ISPRS Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVIII-3/W8:329–333*.
- Hua, B.-S., M.-K. Tran and S.-K. Yeung. 2018. Pointwise convolutional neural networks. Pages 984–993 in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Huang, H., E. Kalogerakis, S. Chaudhuri, D. Ceylan, V. G. Kim and E. Yumer. 2018. Learning local shape descriptors from part correspondences with multi-view convolutional networks. *ACM Transactions on Graphics (TOG)* 37(1):6.
- Kada, M. and A. Wichmann. 2012. Sub-surface growing and boundary generalization for 3D building reconstruction. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences I-3:233–238*.
- Kada, M. and A. Wichmann. 2013. Feature-driven 3D building modeling using planar halfspaces. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences II-3/W3:37–42*.
- Kalogerakis, E., M. Averkiou, S. Maji and S. Chaudhuri. 2017. 3D shape segmentation with projective convolutional networks. Pages 3779–3788 in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Kingma, D. and J. Ba. 2015. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980v8*.
- Klokov, R. and V. Lempitsky. 2017. Escape from cells: Deep Kd-networks for the recognition of 3D point cloud models. Pages 863–872 in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*.
- Krivzhnevsky, A. 2009. Learning multiple layers of features from tiny images. *Technical Report, Department of Computer Science, University of Toronto*.
- Krizhevsky, A., I. Sutskever and G. Hinton. 2012. ImageNet classification with deep convolutional neural networks. In *Proceedings of the 26th International Conference on Neural Information Processing Systems (NIPS)*.
- Li, Y., S. Pirk, H. Su, C. R. Qi and L. J. Guibas. 2016. FPNN: Field probing neural networks for 3D data. In *Proceedings of the 30th Conference on Neural Information Processing Systems (NIPS)*.
- Maturana, D. and S. Scherer. 2015. VoxNet: A 3D convolutional neural network for real-time object recognition. Pages 922–928 in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.
- Munoz, D., J. A. Bagnell, N. Vandapel and M. Hebert. 2009. Contextual classification with functional max-margin Markov networks. Pages 975–982 in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Qi, C. R., H. Su, K. Mo and L. J. Guibas. 2017a. PointNet: Deep learning on point sets for 3D classification and segmentation. Pages 77–85 in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Qi, C. R., H. Su, M. Nießner, A. Dai, M., Yan and L. J. Guibas. 2016. Volumetric and multi-view CNNs for object classification on 3D data. Pages 5648–5656 in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Qi, C. R., L. Yi, H. Su and L. J. Guibas. 2017b. PointNet++: Deep hierarchical feature learning on point sets in a metric space. In *Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS)*.
- Riegler, G., A. O. Ulusoy and A. Geiger. 2017. OctNet: Learning deep 3D representations at high resolutions. Pages 6620–6629 in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Sabour, S., N. Frosst and G. E. Hinton. 2017. Dynamic routing between capsules. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS)*.
- Serna, A., B. Marcotegui, F. Goulette and J.-E. Deschaud. 2014. Paris-Rue-Madame database: A 3D mobile laser scanner dataset for benchmarking urban detection, segmentation and classification methods. Pages 819–824 in *Proceedings of the 3rd International Conference on Pattern Recognition, Applications and Methods (ICPRAM)*.
- Simonyan, K. and A. Zisserman. 2015. Very deep convolutional networks for large-scale image recognition. In *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*.
- Song, S. and J. Xiao. 2014. Sliding shapes for 3D object detection in depth images. Pages 634–651 in *Computer Vision—ECCV*. Edited by D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars. Lecture Notes in Computer Science vol. 8694. Cham, Switzerland: Springer.
- Song, S. and J. Xiao. 2016. Deep sliding shapes for Amodal 3D object detection in RGB-D images. Pages 808–816 in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Su, H., S. Maji, E. Kalogerakis and E. Learned-Miller. 2015. Multi-view convolutional neural networks for 3D shape recognition. Pages 945–953 in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*.
- Szegedy, C., W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich. 2015. Going deeper with convolutions. Pages 1–9 in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Tchapmi, L. P., C. B. Choy, I. Armeni, J. Gwak and S. Savarese. 2017. SEGCloud: Semantic segmentation of 3D point clouds. In *Proceedings of the 5th International Conference on 3D Vision (3DV)*.

- Thrun, S. 2001. Learning occupancy grids with forward models. Pages 1676–1681 in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. Expanding the Societal Role of Robotics in the Next Millennium (Cat. No.01CH37180)* vol. 3.
- Vallet, B., M. Brédif, A. Serna, B. Marcotegui and N. Paparoditis. 2015. TerraMobilita/iQmulus urban point cloud analysis benchmark. *Computers & Graphics* 49:126–133.
- Vosselman, G. and R. Klein. 2010. Visualisation and structuring of point clouds. In *Airborne and Terrestrial Laser Scanning*, edited by G. Vosselman and H.-G. Maas, (eds.), 45–82. Dunbeath, Scotland: Whittles Publishing.
- Wang, R. 2013. 3D building modeling using images and LiDAR: A review. *International Journal of Image and Data Fusion* 4(4):273–292.
- Wang, D. Z. and I. Posner. 2015. Voting for voting in online point cloud object detection. In *Proceeding of the 11th International Conference on Robotics: Science and Systems (RSS)*.
- Weinmann, M., B. Jutzi, S. Hinz and C. Mallet. 2015. Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers. Pages 286–304 in *ISPRS Journal of Photogrammetry and Remote Sensing* vol. 105.
- Wichmann, A., and M. Kada. 2014. 3D building adjustment using planar half-space regularities. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* II-3:189–196.
- Wu, Z., S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang and J. Xiao. 2015. 3D ShapeNets: A deep representation for volumetric shapes. Pages 1912–1920 in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Wu, J., C. Zhang, T. Xue, W. T. Freeman and J. B. Tenenbaum. 2016. Learning a probabilistic latent space of object shapes via 3D generative-adversarial modeling. In *Proceedings of the 30th International Conference on Neural Information Processing Systems (NIPS)*.
- Yi, L., L. Shao, M. Savva, H. Huang, Y. Zhou, Q. Wang, B. Graham, M. Engelcke, R. Klokov, V. Lempitsky, Y. Gan, P. Wang, K. Liu, F. Yu, P. Shui, B. Hu, Y. Zhang, Y. Li, R. Bu, M. Sun, W. Wu, M. Jeong, J. Choi, C. Kim, A. Geethachandra, N. Murthy, B. Ramu, B. Manda, M. Ramanathan, G. P. P. Kumar, S. Srivastava, S. Bhugra B. Lall, C. Häne, S. Tulsiani, J. Malik, J. Lafer, R. Jones, S. Li, J. Lu, S. Jin, J. Yu, Q. Huang, E. Kalogerakis, S. Savarese, P. Hanrahan, T. Funkhouser, H. Su and L. Guibas. 2017. Large-scale 3D shape reconstruction and segmentation from ShapeNet Core55. *arXiv preprint arXiv:1710.06104v2*.
- Zhou, Y. and O. Tuzel. 2018. VoxelNet: End-to-end learning for point cloud based 3D object detection. Pages 4490–4499 in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

WHO'S WHO IN ASPRS

BOARD OF DIRECTORS

BOARD OFFICERS

President

Thomas Jordan, Ph.D.
University of Georgia

President-Elect

Jeff Lovin
Woolpert

Vice President

Jason Stoker
U.S. Geological Survey

Past President

Anne Hillyer, CP
Bonneville Power Administration (USDOE)

Treasurer

Stewart Walker, Ph.D.

Secretary

Roberta E. Lenczowski

BOARD MEMBERS

Corporate Members Council – 2019

Chair: Joe Cantz
Vice Chair: TBA
www.asprs.org/About-Us/Sustaining-Members-Council.html

Early-Career Professionals Council – 2020

Chair: Amanda Aragon
Vice Chair: Peng Fu

Region Officers Council – 2020

Chair: Lorraine B. Amenda, PLS, CP
Vice Chair: Demetrio Zourarakas

Student Advisory Council – 2020

Chair: Victoria Scholl
Vice Chair: TBA
<http://www.asprs.org/Students/Student-Advisory-Council.html>

Technical Division Directors Council – 2020

Chair: John McCombs
Vice Chair: Bandana Kar, Ph.D.

TECHNICAL DIVISION OFFICERS

Geographic Information Systems Division – 2021

Director: Xan Fredericks
Assistant Director: Denise G. Theunissen
www.asprs.org/Divisions/GIS-Division.html

Lidar Division – 2020

Director: Amar K. Nayegandhi, CP, CMS
Assistant Director: Joshua Nimetz, CMS
www.asprs.org/Divisions/Lidar-Division.html

Photogrammetric Applications Division – 2020

Director: Paul C. Bresnahan
Assistant Director: Kurt Rogers
www.asprs.org/Divisions/Photogrammetric-Applications-Division.html

Primary Data Acquisition Division – 2021

Director: Jon Christopherson
Assistant Director: J. Chris Ogier
www.asprs.org/Divisions/Primary-Data-Aquisition-Division.html

Professional Practice Division – 2022

Director: Harold W. Rempel, III, CP
Assistant Director: TBA
www.asprs.org/Divisions/Professional-Practice-Division.html

Remote Sensing Applications Division – 2020

Director: David W. Kreighbaum
Assistant Director: Raechel A. White, Ph.D.
www.asprs.org/Divisions/Remote-Sensing-Applications-Division.html

Unmanned Autonomous Systems (UAS) – 2020

Director: Benjamin Vander Jagt, Ph.D.
Assistant Director: Megan Miller, Ph.D.

REGION PRESIDENTS

Alaska Region

Caitlin Verlund
<http://www.asprsalaska.org/>

Cascades Region

Camile Westlake
<http://columbia.asprs.org/CRR/>

Central New York Region

John Boland, C.P.
<http://www.esf.edu/asprs/>

Eastern Great Lakes Region

Srinivasan Dharmapuri, CP, CMS
<http://egl.asprs.org/>

Florida Region

Erin McCormick
<http://florida.asprs.org/>

Heartland Region

Paul Manley
<http://heartland.asprs.org/>

Intermountain Region

Sowmya Selvarajan, Ph.D.
<http://asprsintermountain.weebly.com/>

Mid-South Region

Demetrio Zourarakis, Ph.D.
<https://www.asprs.org/all-regions/mid-south.html>

New England Region

TBA
<http://neweng.asprs.org/>

North Atlantic Region

Richard W. Carlson, Jr., P.L.S., C.P.
<http://natlantic.asprs.org/>

Pacific Southwest Region

Lorraine B. Amenda, PLS, CP
<https://pswasprs.org/>

Potomac Region

Gang Chen, Ph.D.
<http://www.asprspotomac.org/>

Rocky Mountain Region

Jeffrey Young
<http://www.asprs-rmr.org/>

Western Great Lakes Region

Kirk Contrucci, CP
<http://wgl.asprs.org/>

Through-Water Dense Image Matching for Shallow Water Bathymetry

Gottfried Mandlbürger

Abstract

The introduction of Dense Image Matching (DIM) has reactivated the interest in photogrammetric surface mapping, as it allows the derivation of Digital Elevation Models with a spatial resolution in the range of the ground sampling distance of the aerial images. While the primary field of application is wide-area mapping of topography and urban scenes, charting bathymetry of clear and shallow water areas is equally feasible via application of multimedia photogrammetry. The article specifically investigates the potential of through-water DIM for high resolution mapping of generally low textured shallow water areas using modern techniques like semiglobal matching and off-the-shelf software. In a case study, the DIM-derived underwater surfaces of coastal and inland water bodies are compared to concurrently acquired laser bathymetry data. With an achieved penetration depth of more than 5 m and deviations in the dm-range compared to the laser data as reference, the results confirm the general feasibility of through-water DIM. However, sufficient bottom texture and favorable environmental conditions are a precondition for achieving accurate results.

Introduction

Although topo-bathymetric Light Detection and Ranging (LiDAR) has become the state-of-the-art for mapping shallow coastal and inland water bodies, most modern airborne sensors feature both high resolution multispectral cameras and one or multiple LiDAR channels (Mandlbürger 2018b; Toschi et al. 2018). The multispectral image content can directly be used to estimate water depths by establishing a color-to-depth relation based on physical and/or regression models (Legleiter 2016; Legleiter et al. 2009; Lyzenga et al. 2006). In addition, the rise of Dense Image Matching (DIM) (Remondino et al. 2014) providing a height value for every image pixel via techniques like semiglobal matching (Hirschmüller 2008), has revived the interest in through-water photogrammetry from terrestrial, manned, as well as unmanned airborne and spaceborne platforms for charting sufficiently transparent water bodies (Butler et al. 2002; Dietrich 2016; Hodúl et al. 2018; Westaway et al. 2001). The potential applications include hydromorphology, archaeology, ecology, hydraulics, and the like. Depth estimation via multimedia photogrammetry is generally hampered by progressive blurring due to arbitrary forward and backward scattering of the light rays along their path in the water column (Guenther et al. 2000; Lyzenga 1978a). Despite the accuracy limitations of the photogrammetric approach depending on water depth, water surface planarity, water clarity, etc. (Maas 2015), the availability of imagery with Ground Sampling Distances (GSD) less than 10 cm fuels the hope for high resolution mapping of the shallow water-land-transition zone.

University of Stuttgart, Institute for Photogrammetry, Geschwister-Scholl-Str. 24D, D-70174 Stuttgart (gottfried.mandlbuerger@ifp.uni-stuttgart.de) and TU Vienna, Department of Geodesy and Geoinformation, Gusshausstr. 27-29, A-1040 Vienna (gottfried.mandlbuerger@geo.tuwien.ac.at).

Related Work

For capturing submerged environments with optical remote sensing, the sensor can be either underwater or in the air above the water table. Underwater techniques in general (Massot-Campos and Oliver-Codina 2015) and underwater photogrammetry in particular (Maas 2015) benefit from a short sensor-to-target distance and are therefore suitable for studying complex three-dimensional (3D) structures like coral reefs, submerged ship wrecks, and underwater infrastructure. For efficient area-wide mapping, however, air- and spaceborne techniques are more appropriate (Burns and Delparte 2017; Dietrich 2016; Hodúl et al. 2018).

The main problem in both airborne LiDAR bathymetry and through-water photogrammetry is the refraction at the air-water interface, which causes a change of direction of the image and laser ray (Kotowski 1988; Maas 2015; Westfeld et al. 2017; Yang et al. 2017) as well as a change of the propagation velocity of the laser pulse (Birkebak et al. 2018). In both cases, knowledge of exact water level heights is a prerequisite for precise refraction and run-time correction of the raw measurements. The advantage of LiDAR bathymetry over multimedia photogrammetry is two-fold. On the one hand, LiDAR is a polar data acquisition strategy based on round-trip runtime measurement, i.e., a single measurement is sufficient to obtain a 3D point. On the other hand, the employed green laser light interacts with the water surface, the water column, and the water bottom and, thus, simultaneously delivers height estimates of the surface and the bottom (Guenther et al. 2000; Schwarz et al. 2017). While image-based water level estimation remains a challenging task due to specular reflection at the interface and the dynamic nature of the surface (Rupnik et al. 2015), especially the infrared channel of multispectral images can be used to detect the water-land-boundary (Frazier and Page 2000). At least for standing water bodies, the 3D shoreline enables the derivation of a continuous water surface model in the absence of extensively measured water level heights.

As stated above, depth estimation via multimedia photogrammetry suffers from image ray bending at the air-water interface (Kotowski 1988; Maas 2015; Murase et al. 2008), which causes a violation of the collinearity condition constituting the inherent mathematical model of most photogrammetric software (Förstner and Wrobel 2016; Kraus 2007). This requires a more generic ray path model and sophisticated data processing procedures for bundle block adjustment (Mulsow 2010). Moreover, image blurring due to multi-directional scattering within the water column (Lyzenga 1978b) further hampers image matching and results in a progressive loss of image texture at the water bottom with increasing water depth.

Especially in this respect, DIM could be of advantage with its inherent property (i) of being able to bridge poorly

Photogrammetric Engineering & Remote Sensing
Vol. 85, No. 6, June 2019, pp. 445–454.
0099-1112/19/445–454

© 2019 American Society for Photogrammetry
and Remote Sensing
doi: 10.14358/PERS.85.6.445

textured areas and (ii) of providing good height accuracy also in smooth, low-textured surface parts (Ressl et al. 2016). While accuracy improvements in multimedia photogrammetry exploiting multi-view stereo has already been reported in literature based on sparse feature points (Wolff and Förstner 2000), DIM is expected to share the same potential for high resolution 3D reconstruction of submerged topography.

Research Aims

Against this background, the main objectives of this paper are to evaluate the applicability and the accuracy potential of multimedia through-water DIM for mapping shallow water bathymetry based on off-the-shelf software and processing pipelines for water bodies with different characteristics. Special emphasis is laid on the very shallow zone, where conventional LiDAR bathymetry is not able to distinguish echoes from the air-water interface and the submerged bottom due to the finite pulse length (Legleiter et al. 2016). Especially in areas with water depths < 1 m, through-water DIM might contribute to bridging the gap in the littoral zone between water and land with a point density comparable to topo-bathymetric LiDAR (i.e., 10–20 points/m²). The study also includes an assessment of the potential of multi-view stereo to improve accuracy and reliability of water bottom mapping, which was already studied for feature-based image matching (Wolff and Förstner 2000) but has not been evaluated for DIM so far to the best of my knowledge.

For DIM, the SURE software was chosen as the tool emerged from a scientific background (Wenzel et al. 2013) and the underlying semiglobal matching based concepts are well described in subject literature (Haala and Rothermel 2012; Hirschmuller 2008). Furthermore, SURE provides convenient command-line and library interfaces for conducting the envisaged scientific experiments with full parameter control in contrast to popular Structure-from-Motion tools offering black-box DIM functionality like PhotoScan (Agisoft 2018) or Pix4Dmapper (Pix4D 2018). And finally, the quality of the SURE point clouds and DEMs was confirmed by several independent investigations compared to competing image matching tools as well as LiDAR as reference (Alidoost and Arefi 2017; Remondino et al. 2014; Ressl et al. 2016). The investigated study areas comprise a coastal scene at the German Baltic Sea near Poel, Germany (Song et al. 2015) and a pre-Alpine lake located in the floodplain of the Lech River near Augsburg, Germany (Mandlbürger et al. 2018). Being aware of the accuracy restrictions reported in literature (Maas 2015), the presented case study aims to answer the question under which conditions reliable results can be expected from through-water DIM under real-world conditions. This is addressed by comparing the photogrammetrically derived depth estimates with airborne topo-bathymetric LiDAR data as reference.

The remainder of this article is structured as follows: the section “Materials and Methods” introduces the study areas and details the general processing strategies. In the section “Results and Discussions”, the achieved results are presented for each test site separately and the outcomes are discussed qualitatively by visual inspection and quantitatively by comparing the through-water DIM point clouds and the derived Digital Elevation Models (DEM) thereof with concurrently acquired topo-bathymetric LiDAR data. The concluding section summarizes the findings together with an outlook on future research on subject matters.

Materials and Methods

In this section, the study areas are introduced (section “Study Area and Datasets”) and the applied data processing strategies and evaluation methods are presented (section “Data Processing”).

Study Area and Datasets

For the study at hand, data from a coastal scene and an inland lake were evaluated in detail. Table 1 summarizes the sensor characteristics of the respective airborne data acquisition. The locations of the study areas are marked in the overview map of Figure 1. Area I is located near Poel (Baltic Sea, Germany) and features bright sandy soil to both sides of a small peninsula with embedded patches of dark sea grass (cf. Figure 3). Data acquisition took place in 2014 in the course of a perennial project evaluating the feasibility of using airborne LiDAR bathymetry to map the near shore areas of the German Baltic Sea (Song et al. 2015). Leica’s HawkEye III sensor was employed in the 2014 campaign for concurrently capturing LiDAR and multispectral image data. The system comprises two green laser channels for mapping deep and shallow bathymetry, respectively, as well as a near infrared (NIR) channel for capturing topography and water level heights. In addition to the laser, the RCD30 multispectral camera is an integral part of this hybrid

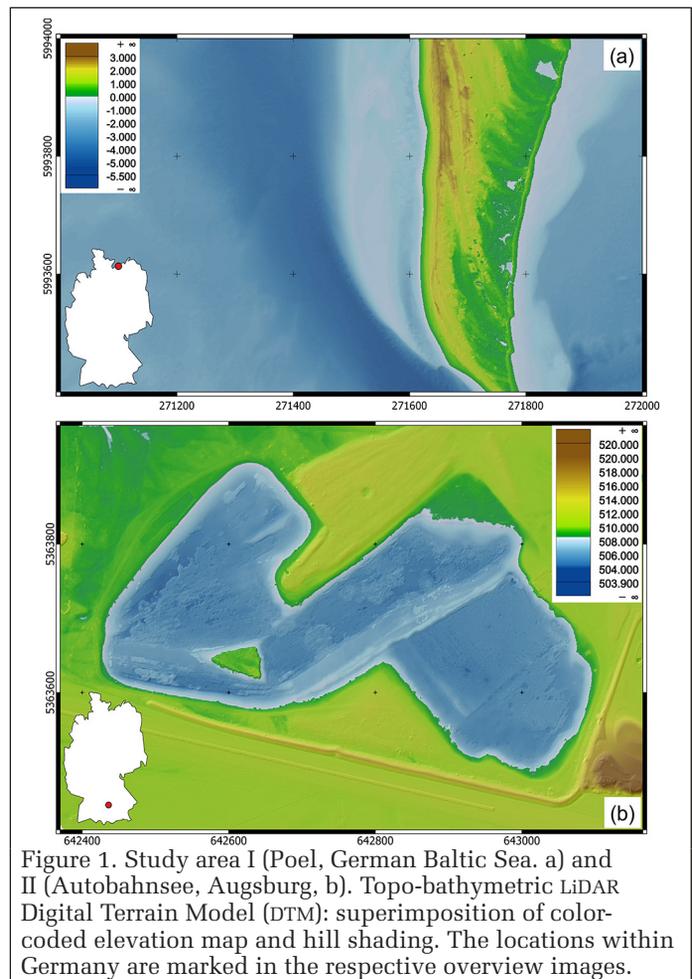


Figure 1. Study area I (Poel, German Baltic Sea, a) and II (Autobahnsee, Augsburg, b). Topo-bathymetric LiDAR Digital Terrain Model (DTM): superimposition of color-coded elevation map and hill shading. The locations within Germany are marked in the respective overview images.

Table 1. Study areas and sensor characteristics.

Area	Site	Year	Location			LiDAR	Camera	Sensor dimension [pixel]	GSD [cm]
			North	East					
I	Poel, Germany	2014	55° 03'	11° 31'		HawkEye III	RCD30	9000 × 6732	6
II	Autobahnsee, Germany	2018	46° 59'	11° 11'		VQ-880-G	DigiCAM 100	11 608 × 8708	5

sensor system and provides 60 MPix Red, Green, Blue, NIR images. Whereas the NIR channel was mainly used for delineating water and land (Frazier and Page 2000), the RCD30 bands in the visible spectrum served as basis for through-water DIM.

Area II (Autobahnsee) is a groundwater supplied lake in the floodplain of the Lech River near Augsburg, Germany (Mandlbürger et al. 2018). The general bottom substrate is bright coarse gravel (grain diameter < 10 cm) with embedded patches of underwater vegetation of different height, density, and color. The bright gravel features a high reflectance in the green domain of the spectrum, which is beneficial for LiDAR bathymetry allowing full coverage of this shallow lake (max. depth: 4.6 m). Occasional patches of algae, in turn, increase the bottom texture, which results in favorable conditions for the application of through-water DIM. Data acquisition took place in April 2018 with a RIEGL VQ-880-G topo-bathymetric laser scanner and two IGI DigiCAM 100 cameras mounted on the same sensor platform. Both cameras are adapted versions of PhaseOne iXU-RS 1000 (100 MPix) sensors. While one of the two cameras recorded Coastal Blue images ($\lambda = 400 - 460$ nm) for a specific experiment (Mandlbürger et al. 2018), the red, green, blue (RGB) imagery of the second constituted the input for the study at hand. Flying at 500–600 m above ground level (a.g.l.) with a speed of 100 knots (50 m/s) yielded aerial images with a GSD of 5 cm and a forward overlap of about 90% (image frame rate: 1.4 Hz). Circular scanning with a constant off-nadir angle of 20° and a laser pulse repetition rate of 550 kHz provided a mean last echo point density of approx. 20 points/m². At the chosen flying height, the diameter of the laser footprint on the ground is around 60 cm (laser beam divergence: 1 mrad). This limits the effective spatial resolution of the topo-bathymetric 3D LiDAR point cloud to approx. 0.5 m.

In both cases, Global Navigation Satellite System (GNSS) and Inertial Measurement Unit (IMU) observations provided aircraft position and attitude information along the entire flight path. Whereas this is mandatory for laser scanning as each laser point has its own exterior orientation, image bundle block adjustment could also rely on precise trajectory information. The latter is of specific importance for open water areas as image tie points are sparse. For both study areas, the captured high-resolution imagery constituted the basis for water depth estimation via multimedia photogrammetry, while LiDAR delivered water surface level data and provided ground truth of the water bottom for evaluating the performance of through-water DIM.

Data Processing

Data processing for each test side consisted of the following processing steps:

- image preprocessing and conversion of the manufacturer-specific raw image files to 16-bit TIFF using the camera software (CaptureOne)
- bundle block adjustment using standard software (Pix4D: Pix4D-mapper, Trimble: Match-AT, Leica: Oriama and IPAS CO+)
- rigorous laser scanning strip adjustment and coregistration of scans and images (Glira et al. 2015; Mandlbürger et al. 2017) using the scientific laser scanning software OPALS (Pfeifer et al. 2014)
- dense image matching using SURE (Rothermel et al. 2012; Wenzel et al. 2013) based on stereo image pairs and multi-image configurations (blocks of up to five consecutive images)
- extraction of water surface laser echoes and modelling of a continuous water surface model
- refraction correction of the raw submerged DIM points (Kotowski 1988; Maas 2015; Wimmer 2016) also considering multi-view stereo configurations

- filtering and smoothing of the refraction corrected through-water DIM points based on standard statistics (i.e., local median, further filtering with Gaussian convolution) and derivation of a topo-bathymetric DEM
- refraction correction of the raw submerged LiDAR points, filtering of ground points (Pfeifer and Mandlbürger 2018), and DTM generation from the classified dry ground and submerged water bottom laser points
- qualitative and quantitative through-water DIM accuracy assessment based on a comparison of the submerged part of the LiDAR DTM and the DIM DEM via visual inspection and statistical analysis (histograms, standard deviation, root mean square error (RMSE), quantiles, etc.)

In the following, some of the processing steps summarized above are commented in more detail. First, proper image orientation via bundle block adjustment (including estimation of the interior orientation, image distortion, and exterior orientation) is a precondition for high quality surface reconstruction via DIM. The inland water dataset (Autobahnsee, Area II) is small enough (600 × 400 m²) so that enough image tie points could be identified in the dry area around the lake (not affected by refraction effects) to properly connect the image block. Multiple overpasses in different flying altitudes (500/600 m a.g.l.) did not only increase the redundancy, but especially the flight lines in higher altitudes helped to stabilize the entire image block. The respective reprojection error of the automatic image tie points for Area II was 0.1 pixel and the RMSE at 10 checkerboard control points, measured in the field with real-time kinematic GNSS, was < 2.5 cm (0.5 pixel) in all three coordinate directions.

For the coastal dataset (Area I), a different strategy was necessary as the predominant part of the aerial images contain water areas only (cf. Figure 1a). In this case, the data provider's bundle block adjustment software (Oriama, IPAS CO+) was employed to estimate the inner and outer orientation of the cameras as well as the mounting calibration parameters (camera lever arms and boresight alignment) based on the dry land parts of the flight block. The final image orientations were subsequently computed by combining the calibrated lever arms and boresight angles with the flight trajectory data calculated from GNSS and IMU observations. The quality of the orientation data was checked by analyzing the DIM results obtained from stereo models on dry ground. As no layering effects could be observed, the image orientations were considered adequate for the purpose of through-water dense matching.

For Area I, the overall registration precision of the LiDAR and DIM point clouds was assessed by deriving DEM rasters with a grid spacing of 25 cm for both data sources independently and by evaluating the height discrepancies. The vertical deviations calculated for smooth dry land areas were below 3 cm confirming a good relative orientation of the image and LiDAR derived point clouds. For Area II, mutual information from concurrently acquired image and scan data constituted the basis for a hybrid scan strip and image bundle block adjustment (Mandlbürger et al. 2017). It is noted here that good relative image orientation is as important as overall absolute orientation because most dense image matching techniques operate on epipolar images to reduce the matching complexity (i.e., matching is only carried out within the epipolar line). Thus, precise relative orientation is a precondition.

Dense matching was basically conducted for stereo image pairs (i.e., no multi-view stereo triangulation) in order to guarantee proper refraction correction of the resulting 3D points according to Snells' law (Kotowski 1988; Maas 2015; Wimmer 2016). As the forward image overlap was more than 60% for all datasets, stereo models were built based on pairs of consecutive images and pairs with one but the next image. The Area II dataset even featured a much higher along track

overlap of approx. 90%. As SURE implements multi-view stereo matching, i.e., the image ray from a certain master image is matched with the homologous points of multiple neighboring images, the resulting overdetermined forward intersection provides additional information for increasing both robustness (outlier detection) and precision (redundancy). This might be especially useful in the bathymetric case as individual ray refraction at the dynamic water surface case causes a violation of the underlying collinearity equation in general and the epipolar geometry in particular. Therefore, the high forward overlap of the Area II dataset was utilized to test multi-view stereo through-water DIM. Figure 2 schematically illustrates ray refraction for the multi-view stereo case.

The correction procedure starts with intersecting the connection line between camera projection center and apparent intersection point P' with the LiDAR derived water surface model. For each image ray, this results in the intersection point at the water surface. Subsequently, the underwater parts of the image rays are refracted at the air-water-interface according to Snell's law (Kotowski 1988). The bent rays intersect at point P . It is noted here that the refracted image rays are skewed and P is therefore calculated within a least squares adjustment as the point minimizing the perpendicular distances to the respective rays.

As can be seen from the principle sketch, proper refraction modelling corrects the inherent water depth underestimation of the apparent ray intersections. Figure 2, however, shows the simplified scenario in a vertical section. In general, the image ray exhibits an arbitrary orientation in space depending on the attitude of the sensor plane and the image coordinates of the respective point on the water bottom. Snell's law of refraction is applicable in the plane containing the image ray and the water surface normal direction. For the laser scanning point cloud, an analogue procedure applies, which, in addition, needs to consider the different propagations velocities in air and water (i.e., run time correction). The scientific laser scanning and point cloud software OPALS (Pfeifer et al. 2014) provides an implementation for both scenarios and was used for the study at hand.

The remaining postprocessing steps comprised the calculation of precision measures for the DIM point clouds (local roughness) at land and within water, and DEM interpolation including smoothing and filtering of measurement errors. Whereas robust interpolation was used for the LiDAR DTM generation including potential detection of water bottom points underneath submerged vegetation (Pfeifer and Mandlbürger 2018), median based filtering was applied for the DIM point clouds. Depending on the dense matching point density, in a first step a 25 cm raster was calculated using the median elevation of all cell points as representative cell value. In most cases further smoothing (Gaussian convolution with a 1×1 m² kernel area) was necessary to obtain acceptable results, thus, reducing the effective resolution of the DIM DEM. Finally, DEM of Differences (DoD) between image- and laser-derived DEMs were calculated to assess the accuracy of dense through-water image matching. The LiDAR dataset acted as reference. Using LiDAR as the comparison basis is justified, as the technique is well established for mapping shallow water areas (Guenther et al. 2000; Mandlbürger et al. 2015). Furthermore, all used sensors (Area I: HawkEye III, Area II: VQ-880-G) satisfy the most rigorous hydrographic survey standards published by the International Hydrographic Organization (IHO) requesting a Total Vertical Uncertainty of less than 25 cm (IHO 2008). The latter specifically applies to those areas providing enough water transparency to allow depth estimation via multimedia photogrammetry.

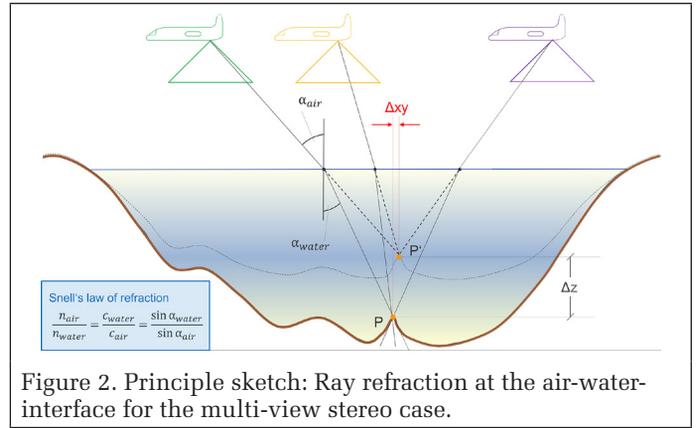


Figure 2. Principle sketch: Ray refraction at the air-water-interface for the multi-view stereo case.

Results and Discussions

In this section, the through-water DIM results are presented and discussed. The overall quality of the achieved DIM DEMs is evaluated by visual inspection and the accuracy is assessed quantitatively based on the topo-bathymetric LiDAR data as reference.

Area I

The composite Figure 3 depicts selected data processing results for a section located in the central part of the Poel dataset (extent: 1 km E-W, 100 m N-S). Figure 3a shows the colored DIM point cloud in ground plan view, highlighting the evident advantage of image-based shallow water mapping, namely the availability of spectral information for each matched point. Figures 3b and 3c depict the DEMs derived from LiDAR and through-water DIM, respectively. Both color-coded elevation maps use blue color tones for the submerged and green-to-brown tones for the dry part of the scene. The LiDAR derived DEM raster appears smooth and reliable to both sides of the water table, proving the high measurement quality of topo-bathymetric LiDAR data (Niemeyer and Soergel 2013; Song et al. 2015). The DIM DEM, in turn, is much rougher in the submerged area. A good height correspondence is achieved at the small peninsula in the eastern part of the plotted area and in the near-shore shallow water areas. Especially the beach area at the western side of the peninsula shows a good DEM height agreement.

This can clearly be seen in the color-coded DoD map plotted in Figure 3d, from which the large systematically positive height differences (> 2 m) in the middle of the investigated area are standing out. Through-water DIM fails entirely at this sandy under water ridge with implausible heights above the water surface due to a lack of texture (smooth sand), further aggravated by a rough water surface. The shoal hereby acts as a natural wave-breaker and causes small overtopping riffles. Bathymetric LiDAR is much less affected in this case, as the laser footprint is typically larger than the small water waves present in this section of the scene and the LiDAR measurement does not depend on the availability of texture (i.e., radiometric differences).

However, the DoD map also reveals that through-water dense matching works quite well in the adjacent deep part of the section with water depths > 4 m. The sea floor is here covered with varying patches of underwater vegetation, which generate texture in the aerial images. In case of low vegetation not affected by movements due to shear stress, this can be beneficial for image matching, whereas high underwater vegetation (eelgrass, etc.) can also hamper DIM, as photogrammetric 3D reconstruction requires multiple views on the objects. A part of the height deviations in this area, visible in the color-coded DoD map (Figure 3d) as small blue or red patches, can therefore be explained by fluctuation of the DIM surface between vegetation canopy and bottom surface, whereas only the bottom is mapped in the LiDAR DTM. However, independent

and concurrently acquired reference data of underwater vegetation would be necessary to answer this question reliably. While the DIM-derived heights show a good agreement with the LiDAR data towards the western boundary of the study area by tendency, local differences in the 1 m-range can still be observed. These differences might again be attributed to the existence of underwater vegetation, but DIM measurement errors due to the wavy water surface are a more likely explanation.

Figure 3e shows the color-coded RMSE roughness map of the DIM dataset. This map allows a quantitative assessment of the dense matching precision as, except at surface discontinuities, the influence of terrain surface roughness is negligible compared to the spread of the measured heights. The region of the peninsula and the adjacent shallow beach area show RMSE values well below 50 cm corresponding to the blue color tones in Figure 3e. The red color, in contrast, indicates a local height spread of more than 2 m, and therefore reveals areas where through-water dense matching is unfeasible for high-resolution description of submerged surfaces.

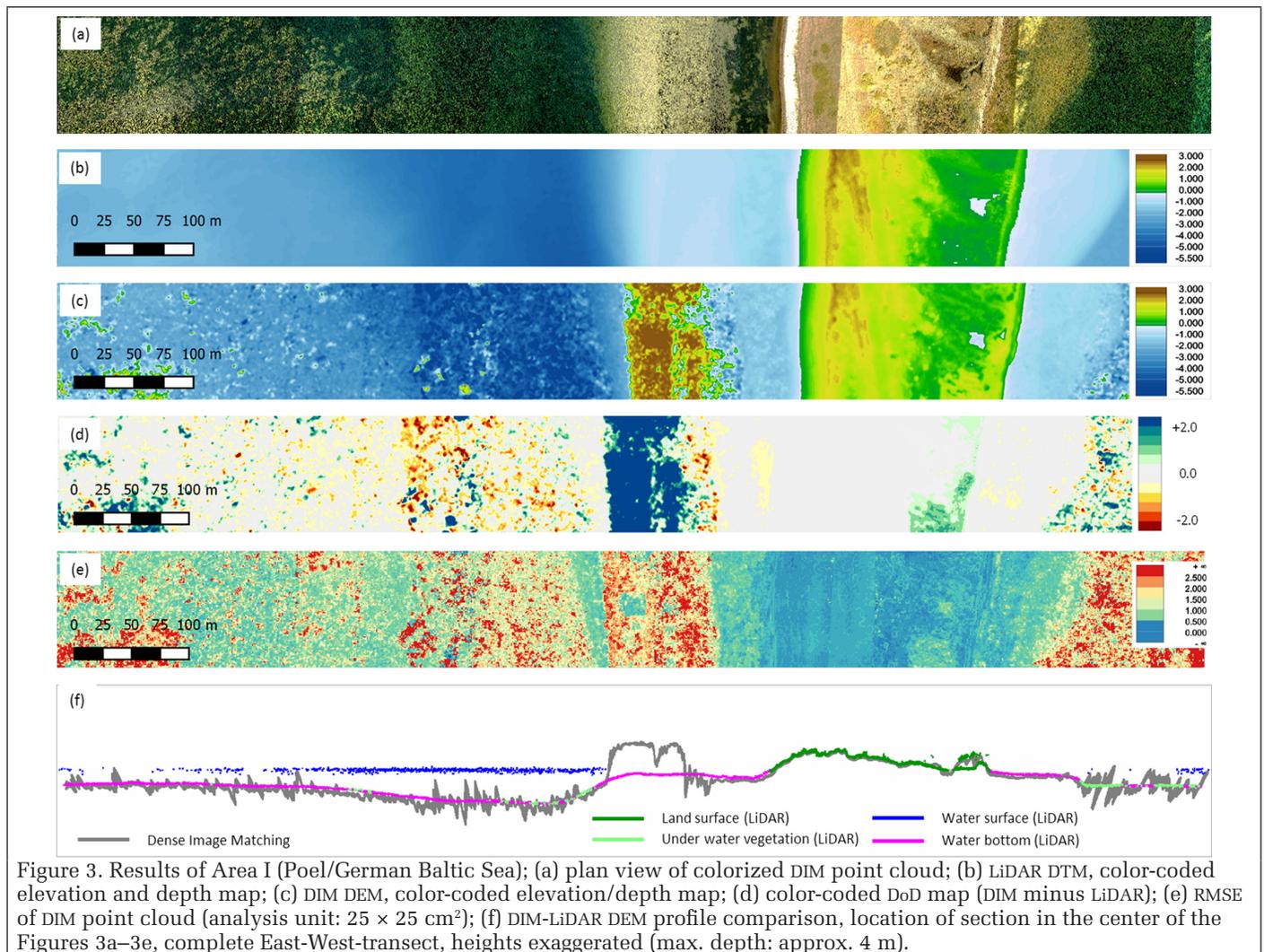
Even if the local precision measures are unfavorable for DIM-based underwater surface mapping in many places, the section view in Figure 3f shows a good general trend between DIM- and LiDAR-derived depths. This even applies to the deeper areas where one might expect inaccurate matching results due to the progressive blurring of submerged terrain features with increasing water depth. However, the general benefit of the higher spatial resolution of pixel-wise dense matching, depending mainly on the GSD of the imagery, compared to

the relatively large laser footprint used in laser bathymetry is lost as large-scale smoothing of the underwater DIM point clouds is inevitable for obtaining acceptable results.

The quantitative comparison for Area I yields a precision (mean local RMSE) of 0.51 m for the through-water dense matching point cloud, and 0.05 m for the LiDAR dataset. Thus, for the entire Area I, the local spread of the height values is 10-times larger for DIM compared to bathymetric LiDAR. The same calculation carried out for dry land (i.e., peninsula) only results in a RMSE of 0.15 m for DIM and 0.07 m for LiDAR, respectively, thus only a factor of two. The decline of precision in this open water scenario amounts to a factor of five. Maas (2015) reports an accuracy decrease of a factor of two under laboratory conditions (planar water surface due to glass interface, clear water, etc.). The additional accuracy drop factor of 2.5 for Area I seems plausible as the open water case at hand exhibits more incalculable factors, from which the currently limited possibilities for capturing the spatio-temporally transient water surface is considered the most prominent one.

Area II

The through-water DIM DEM for Area I resulted from pairwise matching. In addition, multi-view stereo (i.e., multi-stereo forward intersection) was tested for the inland lake dataset of Area II using blocks of five consecutive images while iterating through the entire image strip (i.e., first run: images 1–5, second run: images 2–6, etc.). This enables both outlier detection and noise reduction within the DIM procedure and potentially



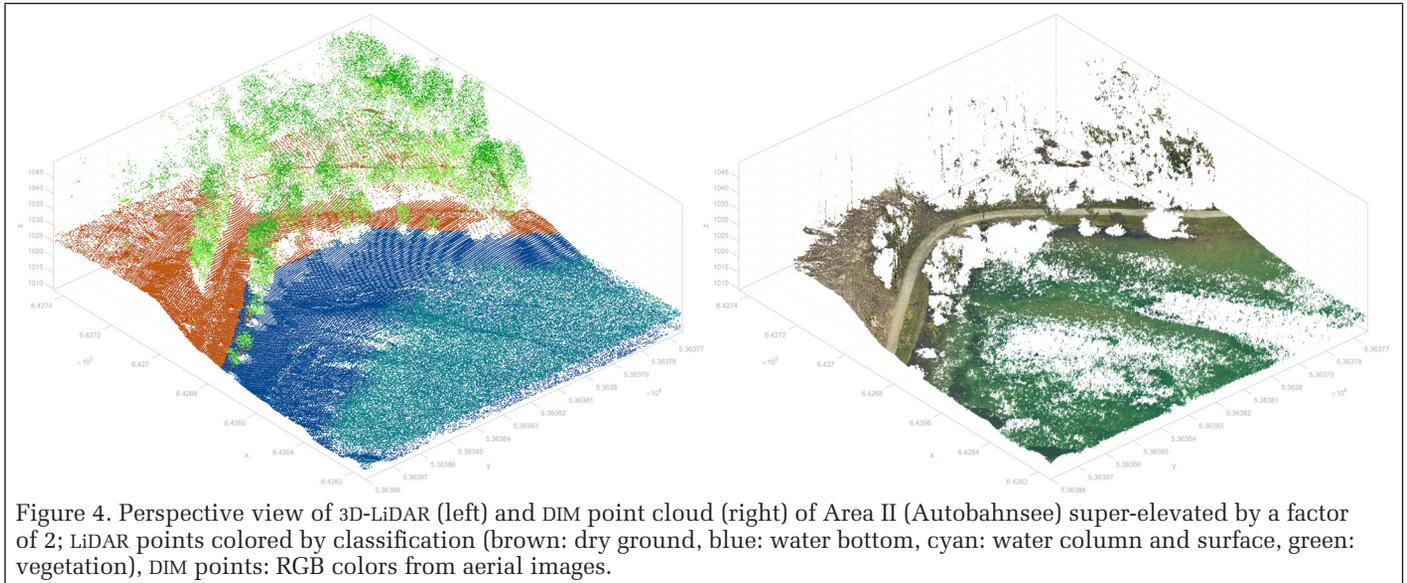


Figure 4. Perspective view of 3D-LiDAR (left) and DIM point cloud (right) of Area II (Autobahnsee) super-elevated by a factor of 2; LiDAR points colored by classification (brown: dry ground, blue: water bottom, cyan: water column and surface, green: vegetation), DIM points: RGB colors from aerial images.

results in data voids, in case the matching results are found to be incoherent at a specific location (Mandlbürger et al. 2017).

Figure 4 shows on the left side a perspective view of the 3D LiDAR points of a single flight line classified into bare ground (brown), water bottom (blue), vegetation (green), and water column (cyan). The right side of Figure 4 depicts the DIM points colored by RGB in the same viewing geometry. The DIM points were derived with multi-stereo forward intersection from a single block of five consecutive images. The location of the scene is marked in the orthophoto of Figure 5b.

The two plots allow a first qualitative assessment: LiDAR is able to penetrate both vegetation and water, and provides continuous coverage of bare earth and submerged water bottom points. Additional laser echoes within the water column stem from volume backscattering. The DIM dataset, in contrast, exhibits data voids beneath trees as well as within the submerged part of the scene. The poor vegetation penetration is a well-known property of 3D point clouds derived from stereo photogrammetry and therefore does not require further discussion. The data voids in the water, in turn, stem from outlier rejection in the multi-view stereo data processing pipeline indicating insufficiently coherent matching results. Whereas the lack of full spatial coverage is a drawback, it also highlights the outlier detection capability of multi-view stereo in general and of the SURE software in particular. However, Figure 4 also exhibits a smooth run of the submerged bottom points for both LiDAR and DIM in open water areas with good bottom texture.

Figure 5 depicts the depth maps derived from through-water DIM and LiDAR, respectively. Again, the LiDAR derived water depth model (Figure 5d) outperforms the DIM results (Figures 5a and 5c) with regard to coverage and consistency. In fact, the LiDAR sensor provided water bottom points down to the maximum depth of the lake (4.6 m) and the run of the submerged surface is smooth. This especially applies to the bare gravel parts of the lake not covered by underwater vegetation. The respective areas are marked with blue polygons in the orthophoto map of Figure 5b, which also emphasizes areas with bright and dark (green) vegetation as well as sun glint areas (orange). Depending on the density and height of the underwater vegetation, the laser echoes partially succeeded in penetrating the vegetation layer. However, especially in the large densely vegetated patch in the western part of the lake, no ground points but rather the canopy was measured.

Figures 5a and 5c show the depth maps derived from pairwise matching and multi-view stereo,

respectively. It is apparent, that pairwise matching delivers a much higher coverage at the prize of a higher volatility of the surface. Multi-view stereo, in turn, only succeeds in the areas without or with only sparse vegetation, but the depth map is much smoother and more reliable compared to the pairwise matching results, which exhibit a chaotic run of the depth contour lines in some vegetated parts of the lake. Figure 5e depicts the precision (i.e., standard deviation of DIM elevations calculated for $25 \times 25 \text{ cm}^2$ cells) of the multi-view stereo DIM point cloud. As can be seen, the height spread within the cells is moderate with dominating green tones in the color-coded roughness map corresponding to a precision of 10 cm with only occasional yellow tones (around 30 cm). Again, this highlights the outlier rejection and smoothing capabilities of multi-view stereo. The color-coded DoD map of Figure 5f, finally, underlines the good overall agreement between LiDAR and DIM derived water depths. Especially the bare soil areas generally show white color tones corresponding to a depth deviation of less than 10 cm. Larger deviations with mostly blue color tones occur in the vegetated part and can be attributed to the limited penetration capability of the photogrammetric approach. Sun glint, in turn, is the most likely reason for the prominent red patch (i.e., DIM depth > LiDAR depth) in the southwestern part of the lake. This underlines the general necessity of sun glint removal as an essential image data preprocessing step (Lyzenga et al. 2006).

Based on the qualitative evaluation described above, a detailed quantitative accuracy assessment of the water depths derived from (multi-view stereo) through-water DIM with regard to LiDAR as reference was carried out for different categories (bare gravel soil, bright and dark vegetation, sun glint). The respective areas of interest polygons are plotted in Figure 5b. Figure 6 shows histograms of the depth deviations for the entire water area (a), the areas with bright vegetation (b), and the bare soil results (c), respectively. Table 2 summarizes the statistics for all categories.

Table 2. Detailed statistics of LiDAR vs. multi-view stereo through-water DIM depth deviations; units: [m].

	# Samples	Mean	Median	σ	σ_{MAD}	q25	q75
total area	339 575	0.086	0.021	0.389	0.118	-0.038	0.162
bare soil	123 772	0.000	-0.005	0.109	0.050	-0.036	0.032
bright vegetation	42 842	0.178	0.146	0.501	0.299	-0.043	0.365
dark vegetation	68 198	0.216	0.190	0.594	0.315	0.001	0.431
sun glint	5378	0.253	0.221	0.460	0.289	0.047	0.440

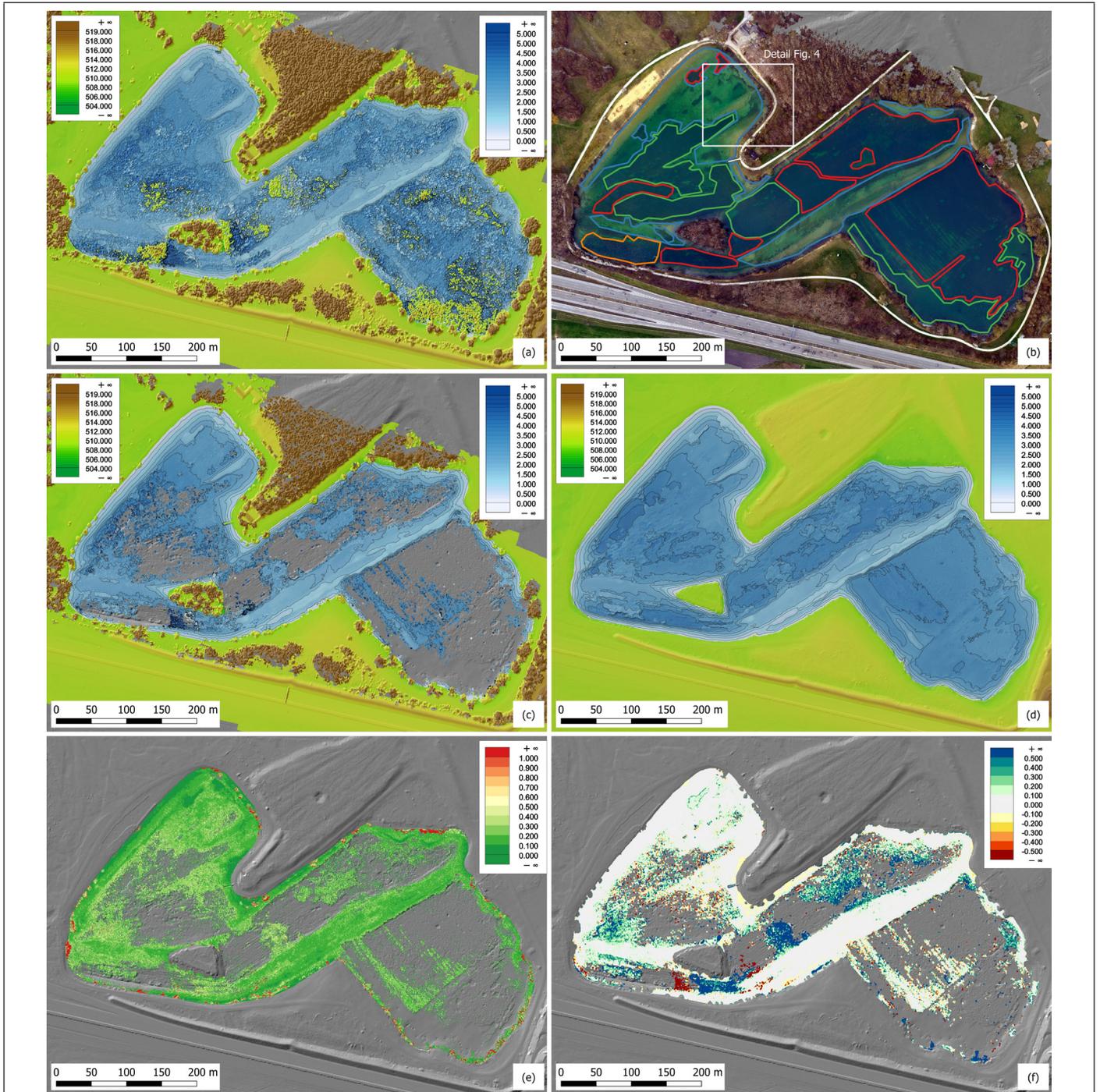


Figure 5. Results of Area II (Autobahnsee, Germany): (a) DIM DEM (pairwise matching, shaded relief map superimposed with color-coded terrain elevations and water depths), (b) RGB Orthophoto, (c) DIM DEM (multi-view stereo matching), (d) LiDAR DTM, (e) multi-view stereo DIM DEM roughness map (local std. dev), (f) DoD LiDAR-DIM (color-coded depth differences).

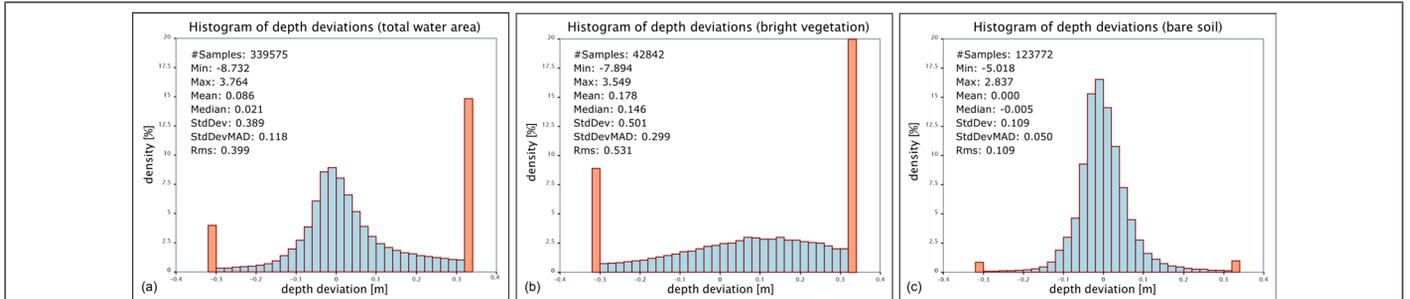


Figure 6. Histograms of LiDAR vs. multi-view stereo through-water DIM depth deviations, (a) total lake area, (b) bright vegetation subarea (bright green polygons in Figure 5b), (c) submerged bare soil subarea (blue polygons in Figure 5b).

Comparing the water depths of DIM and LiDAR for the entire lake area results in an error distribution slightly biased in positive direction (i.e., DIM derived bottom surface too high; mean: 9 cm; median: 2 cm) with a standard deviation (σ) of 39 cm. The bias is mainly a result of the prominent group of large positive deviations > 30 cm caused by the limited penetration capability of photogrammetric surface reconstruction. In the following, robust statistics (median, σ_{MAD}) are used for the further discussion as they are less sensitive to outliers. σ_{MAD} is the Median of Absolute Differences with regard to the median (referred to as *stdDevMAD* in Figure 6) and constitutes a robust estimator of the standard deviation (σ).

Concentrating on the lake part with bright and dark vegetation, the distribution of the depth deviations show a more pronounced bias in the positive direction (median: 15/19 cm) as well as a higher roughness (σ_{MAD} : 30/32 cm) compared to the overall results (σ_{MAD} : 12 cm). The higher values occur in the area with dark vegetation. As expected, low radiometric signal strength has a negative influence on the achievable accuracy. The bright vegetation DoD histogram of Figure 6b further reveals two striking peaks with large positive and negative deviations and an otherwise very flat distribution for deviations in the acceptable range (< 30 cm).

The bare soil area, in turn, shows a very good agreement between LiDAR and DIM derived depths. The error distribution is unbiased (mean: 0.0 cm, median: -0.5 cm) with a low standard deviation (σ : 11 cm, σ_{MAD} : 5 cm) corresponding to 1–2 pixels in the image. As the general height accuracy expectation of DIM in the topographic case is in the range of 0.5–1 pixel depending on the radiometric quality and the accuracy of the image orientations (Ressl et al. 2016), the above results clearly demonstrate the potential of through-water DIM under favorable conditions. The values are also well in line with comparable investigations reported in literature (Maas 2015). The interquartile range further underlines the good symmetry of the deviations in the bare soil class (q25/q75: -3.6/+3.2 cm), whereas this is not the case for the other categories (e.g. dark vegetation: 0.1/43.1 cm).

Conclusions and Outlook

In this contribution, the general applicability of off-the-shelf dense image matching software for mapping shallow water bathymetry was evaluated. Tests were carried out for a coastal dataset (Area I: Baltic Sea, Germany) and a standing inland water body (Area II: Autobahnsee, Augsburg, Germany) using the software SURE (Rothermel et al. 2012; Wenzel et al. 2013), which implements a variant of the Semi Global Matching (SGM) algorithm (Hirschmuller 2008). In general, the feasibility of using through-water DIM for deriving shallow water bathymetry could be verified, but the resulting point clouds are less reliable compared to laser bathymetry, which is still the state-of-the-art for shallow water mapping.

Although further investigations including independent ground truth measurements are necessary to verify the findings of this case study, the following main conclusions can be drawn:

- Through-water dense matching worked well given (i) clear water conditions, (ii) calm water surface, and (iii) availability of sufficient bottom texture. Whereas especially the latter is a general precondition for photogrammetry, the advent of DIM (Hirschmuller 2008) has shown that pixel-wise height estimation is possible even in poorly textured areas due to the additional smoothness constraint within the SGM algorithm.
- While the water bottom surface is usually smooth because of the water's shear stress, neighboring image pixels often show a substantial radiometric variation due the spatio-temporally transient (i.e., wavy) structure of the water surface causing refraction of the submerged part of the image ray in all possible directions. Thus, apparent parallaxes do

not only occur in image base direction but also perpendicular to the flight trajectory. For speeding up processing, the SURE software first performs epipolar image rectification and subsequently implements dense matching only within the epipolar line. While this performance improvement is not part of the original SGM-approach (Hirschmuller 2008), it is a commonly used DIM software technique (e.g. Trimble/Match-T). This, however, hampers feature matching if the image rays are refracted away from the epipolar line. Future investigations will therefore concentrate on quantifying and compensating the effect of arbitrarily tilted water surfaces.

- While aerial images with a GSD of 5–6 cm captured from manned aerial platforms were investigated in the study at hand, multimedia photogrammetry is nowadays also employed from remotely piloted, unmanned platforms (Alidoost and Arefi 2017; Butler et al. 2002) potentially providing GSDs in the cm-range. Smaller pixel sizes, however, will only then lead to an increased spatial resolution of the submerged topography, if capturing and modeling the dynamic water surface is feasible at cm-resolution as well. As stated above, this will be a major topic for future research.
- In clear water areas with sufficient bottom texture, a good correspondence with the bathymetric LiDAR data was observed. An acceptable run of the DIM-based under-water surface, however, could only be achieved by extensive low-pass filtering of the original point cloud. This, in turn, decreases the effective spatial resolution considerably. With an applied convolution kernel size of 1–2 m, the spatial resolution of the DIM DEM effectively drops below the achievable resolution of topo-bathymetric LiDAR. The latter is rather restricted by the diameter of the laser footprint on the ground of typically 50–60 cm than by the point density, which is often in the range of 10–20 points/m².
- An accuracy gain could be observed by using multi-view stereo instead of pairwise matching. The additional complexity of the multi-ray refraction problem was by far compensated by a smoother run of the surface. Using multi-view stereo reduced the necessary additional averaging in post-processing as the over-determined forward intersection also provides smoothing.
- A further advantage of using multi-view stereo is the inherent outlier detection capability, as points are only accepted if the matching results are sufficiently coherent. Using pairwise matching, elimination of unreliable points has to rely on statistical analysis of the point cloud roughness (local standard deviation) in post-processing.
- In further tests, not detailed in this paper, a quality decrease of the submerged DIM point cloud could be observed in case of turbid water conditions (Mandlbürger 2018a). Turbidity evidently reduces the contrast and poor texture directly influences the achievable accuracy.
- The conducted real-world experiments clearly showed that the submerged DIM point clouds and DEMs derived thereof are acceptable in some areas but quickly become unreliable in case that at least one of the preconditions formulated above is not met. The study therefore also investigated methods to identify usable and unreliable areas, respectively. The local height spread (RMSE) has proven to be a good indicator.

To sum it up, the case study confirmed that through-water dense image matching is feasibly for mapping shallow water bathymetry under favorable conditions (clear water, calm water surface, bottom texture), but DIM suffers more from adverse environmental conditions (turbidity, wavy water surface, poor bottom texture, dark image areas) compared to bathymetric LiDAR. The latter profits from the fact of being an active, polar data acquisition technique. A single measurement is, thus, sufficient to record a 3D point if the signal strength of the laser echo reflected from the water bottom is strong enough.

Through-water dense image matching, in turn, might further benefit from simultaneous capture of LiDAR and nadir/oblique images. While this has already been investigated for urban environments (Toschi et al. 2018), comparable studies for mapping bathymetry are still outstanding and respective sensor configurations (i.e., multi-head cameras and green LiDAR) are not available to date.

Apart from this, off-the-shelf DIM software relies on the collinearity equation in general and uses epipolar geometry in particular, but does not consider bending of the image rays due to refraction at the air-water interface. Whereas a general compensation of the apparent light rays was carried out in post-processing after image matching (Pfeifer et al. 2014; Wimmer 2016), light refraction also deteriorates the matching results in the first place. Future work on subject matters will therefore focus on potential corrections in the image space before stereo matching rather than correcting the raw DIM point clouds in post-processing in object space. Whereas this is a promising approach from a theoretical point of view, it still requires precise mapping of the dynamic water surface, which is a further research area. Good prospects of success for a coupled data processing strategy are expected for concurrently acquired bathymetric LiDAR and multispectral image data. A trend towards a widespread availability of such hybrid systems is clearly discernible.

Acknowledgments

The presented study was conducted within the project “Bathymetry by fusion of airborne laser scanning and multispectral aerial imagery” funded by the German Research Foundation (DFG). The author wants to thank Christian Mayr and Hannes Nuebel (undergraduate students at the University of Stuttgart) for their support with data processing of Area II.

References

- Agisoft. 2019. Agisoft PhotoScan. <<http://www.agisoft.com/>> Accessed 3 January 2019.
- Alidoost, F. and H. Arefi. 2017. Comparison of UAS-based photogrammetry software for 3D point cloud generation: A survey over a historical site. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences IV-4/W4:55–61*.
- Birkebak, M., F. Eren, S. Pe'eri and N. Weston. 2018. The effect of surface waves on airborne Lidar bathymetry (ALB) measurement uncertainties. *Remote Sensing* 10(3):453. <<http://www.mdpi.com/2072-4292/10/3/453>>.
- Burns, J. and D. Delparte. 2017. Comparison of commercial structure-from-motion photogrammetry software used for underwater three-dimensional modeling of coral reef environments. Pages 127–131 in *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, held in Nafplio, Greece.
- Butler, J., S. Lane, J. Chandler and E. Porfiri. 2002. Through-water close range digital photogrammetry in flume and field environments. *The Photogrammetric Record* 17(99):419–439. <<https://onlinelibrary.wiley.com/doi/abs/10.1111/0031-868X.00196>>.
- 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>>.
- Förstner, W. and B. P. Wrobel. 2016. *Photogrammetric Computer Vision: Statistics, Geometry, Orientation and Reconstruction*. 643–725. Cham, Switzerland: Springer International Publishing. <http://dx.doi.org/10.1007/978-3-319-11550-4_15>.
- Frazier, P. S. and K. J. Page. 2000. Water body detection and delineation with Landsat TM data. *PE&RS, Photogrammetric Engineering & Remote Sensing* 66(12):1461–1467.
- Glira, P., N. Pfeifer, C. Briese and C. Ressel. 2015. A correspondence framework for ALS strip adjustments based on variants of the ICP algorithm. *PGF Photogrammetrie, Fernerkundung, Geoinformation* 2015(4):275–289. <<http://dx.doi.org/10.1127/pfg/2015/0270>>.
- Guenther, G. C., A. G. Cunningham, P. E. Laroque and D. J. Reid. 2000. Meeting the accuracy challenge in airborne Lidar bathymetry. In *Proceedings of the 20th EARSeL Symposium: Workshop on Lidar Remote Sensing of Land and Sea*, held in Dresden, Germany.
- Haala, N. and M. Rothermel. 2012. Dense multi-stereo matching for high quality digital elevation models. *PGF Photogrammetrie, Fernerkundung, Geoinformation* 2012(4):331–343. <<http://www.ingentaconnect.com/content/schweiz/pfg/2012/00002012/00000004/art00003>>.
- Hirschmuller, H.. 2008. Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30(2):328–341. <<http://dx.doi.org/10.1109/TPAMI.2007.1166>>.
- Hodúl, M., S. Bird, A. Knudby and R. Chénier. 2018. Satellite derived photogrammetric bathymetry. *ISPRS Journal of Photogrammetry and Remote Sensing* 142:268–277. <<http://www.sciencedirect.com/science/article/pii/S0924271618301783>>.
- IHO. 2008. *S-44, Standards for Hydrographic Surveys*. Monaco. <http://www.iho.int/iho_pubs/standard/S-44_5E.pdf>.
- Kotowski, R.. 1988. Phototriangulation in multi-media-photogrammetry. Pages 324–334 in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, held in Kyoto, Japan.
- Kraus, K. 2007. *Photogrammetry—Geometry from Images and Laser Scans*. 2nd ed. Berlin, Germany: De Gruyter.
- Legleiter, C. J. et al. 2016. Evaluating the capabilities of the CASI hyperspectral imaging system and aquarius bathymetric LiDAR for measuring channel morphology in two distinct river environments. *Earth Surface Processes and Landforms* 41(3):344–363.
- Legleiter, C. J. 2016. Inferring river bathymetry via image to depth quantile transformation (IDQT). *Water Resources Research* 52(5):3722–3741. <<https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016WR018730>>.
- Legleiter, C. J., A. R. Dar and L. L. Rick. 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>>.
- Lyzenga, D. R. 1978a. Passive remote-sensing techniques for mapping water depth and bottom features.. *Applied Optics* 17(3):379–383.
- Lyzenga, D. R. 1978b. Passive remote sensing techniques for mapping water depth and bottom features. *Applied Optics* 17(3):379–383.
- Lyzenga, D. R., N. P. Malinas and F. J. Tanis. 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):18140–52. <<http://www.mdpi.com/1424-8220/15/8/18140>>.
- Mandlbürger, G. et al. 2017. Improved topographic models via concurrent airborne LiDAR and dense image matching. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences IV-2/W4:259–266*. <<https://www.isprs-ann-photogramm-remote-sens-spatial-inf-sci.net/IV-2-W4/259/2017/>>.
- Mandlbürger, G. 2018a. A case study on through-water dense image matching. Pages 659–666 in *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, held in Riva del Garda, Italy.
- Mandlbürger, G. 2018b. A review of airborne laser bathymetry sensors. Pages 41–42 in *New Challenges in Hydraulic Research and Engineering*, held in Trento, Italy.

- Mandlbürger, G., C. Hauer, M. Wieser and N. Pfeifer. 2015. Topobathymetric LiDAR for monitoring river morphodynamics and instream habitats—A case study at the Pielach River. *Remote Sensing* 7(5):6160–6195. <<http://www.mdpi.com/2072-4292/7/5/6160>>.
- Mandlbürger, G., J. Kremer, F. Steinbacher and R. Baran. 2018. Investigating the use of coastal blue imagery for bathymetric mapping of inland water bodies. Pages 275–282 in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, held in Karlsruhe, Germany.
- Massot-Campos, M. and G. Oliver-Codina. 2015. Optical sensors and methods for underwater 3D reconstruction. *Sensors (Switzerland)* 15(12):31525–57.
- Mulsow, C.. 2010. A flexible multi-media bundle approach. Pages 472–477 in *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, held in Newcastle upon Tyne, England.
- Murase, T. et al. 2008. A photogrammetric correction procedure for light refraction effects at a two-medium boundary. *Photogrammetric Engineering and Remote Sensing* 74:1129–1136.
- Niemeyer, J. and U. Soergel. 2013. Opportunities of airborne laser bathymetry for the monitoring of the sea bed on the Baltic Sea Coast. Pages 179–184 in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, held in Antalya, Turkey.
- Pfeifer, N. and G. Mandlbürger. 2018. Lidar data filtering and digital terrain model generation. In *Topographic Laser Ranging and Scanning—Principles and Processing*, 2nd ed., Edited by J. Shan and C. K. Toth, 349–378. Boca Raton, Fla.: CRC Press.
- Pfeifer, N., G. Mandlbürger, J. Otepka and W. Karel. 2014. OPALS—A framework for airborne laser scanning data analysis. *Computers, Environment and Urban Systems* 45:125–136. <<http://www.sciencedirect.com/science/article/pii/S0198971513001051>>.
- Pix4D. 2019. *Pix4Dmapper: Professional Drone Mapping and Photogrammetry Software*. <<https://www.pix4d.com/product/pix4dmapper-photogrammetry-software>> Accessed 3 January 2019.
- Remondino, F., M. G. Spera, E. Nocerino, F. Menna and F. Nex. 2014. State of the art in high density image matching. *The Photogrammetric Record* 29(146):144–166. <<http://dx.doi.org/10.1111/phor.12063>>.
- Ressl, C., H. Brockmann, G. Mandlbürger and N. Pfeifer. 2016. Dense image matching vs. airborne laser scanning—Comparison of two methods for deriving terrain models. *PFG Photogrammetrie, Fernerkundung, Geoinformation* 2016(2):57–73.
- Rothermel, M., K. Wenzel, D. Fritsch and N. Haala. 2012. SURE: Photogrammetric surface reconstruction from imagery. In *Proceedings of the Low Cost 3D Workshop, Berlin*, held in Berlin, Germany. <http://www.ifp.uni-stuttgart.de/publications/2012/Rothermel_et_al_lc3d.pdf> Accessed 2 July 2014.
- Rupnik, E., J. Jansa and N. Pfeifer. 2015. Sinusoidal wave estimation using photogrammetry and short video sequences. *Sensors* 15(12):30784–809. <<http://www.mdpi.com/1424-8220/15/12/29828>>.
- Schwarz, R., N. Pfeifer, M. Pfennigbauer and A. Ullrich. 2017. Exponential decomposition with implicit deconvolution of lidar backscatter from the water column. *PFG—Journal of Photogrammetry, Remote Sensing and Geoinformation Science* 85(3):159–167. <<https://doi.org/10.1007/s41064-017-0018-z>>.
- Song, Y. et al. 2015. Comparison of three airborne laser bathymetry data sets for monitoring the German Baltic Sea Coast. *Proceedings of SPIE* 9638:9638-9638–39. <<https://doi.org/10.1117/12.2194960>>.
- Toschi, I., F. Remondino, R. Rothe and K. Klimek. 2018. Combining airborne oblique camera and Lidar sensors : Investigation and new perspectives. Pages 437–444 in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, held in Karlsruhe, Germany.
- Wenzel, K., M. Rothermel, N. Haala and D. Fritsch. 2013. SURE—The IFP software for dense image matching. Pages 59–70 in *Photogrammetric Week '13*. Edited by D. Fritsch. Wichmann/VDE Verlag, Berlin & Offenbach. <<http://www.ifp.uni-stuttgart.de/publications/phowo13/080Wenzel.pdf>> Accessed 27 March 2017.
- Westaway, R. M., S. N. Lane and D. M. Hicks. 2001. Remote sensing of clear-water, shallow, gravel-bed rivers using digital photogrammetry. *Photogrammetric Engineering and Remote Sensing* 67(11):1271–1281.
- Westfeld, P., H. G. Maas, K. Richter and R. Weiß. 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>>.
- Wimmer, M. 2016. Comparison of active and passive optical methods for mapping river bathymetry. *TU Wien*. <<https://resolver.obvsg.at/urn:nbn:at:at-ubtuw:1-2619>>.
- Wolff, K. and W. Förstner. 2000. Exploiting the multi view geometry for automatic surfaces reconstruction using feature based matching in multi media photogrammetry. Pages 900–907 in *International Archives of Photogrammetry and Remote Sensing Sciences*, Amsterdam, the Netherlands.
- Yang, F. et al. 2017. Refraction correction of airborne LiDAR bathymetry based on sea surface profile and ray tracing. *IEEE Transactions on Geoscience and Remote Sensing* 55(11):6141–6149.

Object-Based Point Cloud Analysis for Landslide and Erosion Monitoring

Andreas Mayr, Martin Rutzinger, and Clemens Geitner

Abstract

Today point clouds from close-range sensing are used for operational erosion and landslide monitoring. Distances between points from multi-temporal acquisitions can indicate surface deformation, while a designation of the underlying geomorphological processes is often handicapped by complex terrain structures and vegetation. We present an approach to landslide monitoring that integrates semantic information and three-dimensional deformation detection automatically. Surface changes are assigned to (i) semantic object classes (landslide scarp, eroded area, deposit) and (ii) to spatially contiguous, individual objects (like parts of the landslide scarp and moving clods of turf and soil). We demonstrate this object-based approach with a time series of 13 topographic Light Detection and Ranging point clouds, covering a site affected by shallow landsliding. The results of this case study illustrate how the presented methods translate the unstructured point clouds into information on geomorphological process dynamics to support erosion and landslide assessment.

Introduction

The observation and investigation of mass movements and erosion processes provides important information for natural hazard management and, more specifically, for the protection and management of steep agricultural land in mountain areas (Turner and Schuster 1996; Alewell *et al.* 2015). These processes may endanger infrastructure or cultivated land, thus compromising human economic activities, and even human lives in mountain areas (Kjekstad and Highland 2009; Petley 2012). A major challenge for sustainable agriculture in landslide-prone areas is soil degradation and erosion (Wiegand and Geitner 2010; Alewell *et al.* 2015), which can be caused either directly by landsliding or by secondary erosion processes.

The monitoring of landslides and soil degradation processes delivers important information for decision makers in order to minimize risk and ensure adequate planning of mitigation measures. For this purpose, close-range photogrammetry (Lucieer *et al.* 2014; Stumpf *et al.* 2015; Scaioni *et al.* 2015) and topographic Light Detection and Ranging (LiDAR) (Jaboyedoff *et al.* 2012; Vericat *et al.* 2014; Scaioni *et al.* 2014; Kromer *et al.* 2017) are well established surveying techniques that provide three-dimensional (3D) point clouds in very high resolution. The systematic and automated change analysis in the original 3D point clouds, which represent natural objects as unstructured surfaces, is still a matter of research. Current studies analyze point cloud distances between multi-temporal data sets, which can then be interpreted by geomorphological experts (Lague *et al.* 2013; Stumpf *et al.* 2015; Fey and Wichmann 2017). However, the interpretation of such distances

between epochs (i.e., data from different, sequential acquisitions) is not straightforward, especially if vegetation cover is present, since changes may originate not only from geomorphological activity but also from seasonal and longer-term vegetation development. With growing time series and the emergence of very dense point clouds at increasing spatial extents, for instance from unmanned aerial vehicle (UAV)-based LiDAR or dense image matching (e.g. Glira *et al.* 2016; Cramer *et al.* 2018), automated processing pipelines for environmental and infrastructure monitoring are gaining in importance.

In this article, we extend our previous work (Mayr *et al.* 2017; Mayr *et al.* 2018) on 3D landslide monitoring with object-based point cloud analysis to a time series of 13 terrestrial LiDAR epochs and assess the applicability of the approach. The processing pipeline removes vegetation points and relates surface deformations detected between multi-temporal point clouds directly to (i) semantic object classes (“landslide scarp”, “eroded area”, “deposit”) and (ii) to individual, geomorphologically meaningful objects (like sections of the landslide scarp and dislocated pieces of material within the landslide).

Test Site and Data

The test site (Figure 1) is located in the valley of Schmirn in Tyrol (Austria). It comprises two shallow landslides on a south-west oriented slope at about 1700 m above sea level. The average gradient of the site is 35°. The size of the eroded area of the larger landslide is 20 × 30 m. Reaching a depth of approximately two meters, the landslides and erosive processes affect the soil and Quaternary deposits on top of Bündner schist bedrock. The two landslides are surrounded by pasture land and meadows. Especially the upper part is sparsely covered with larch (*Larix decidua*) and shrubs, mainly green alders (*Alnus alnobetula*). In some places the bedrock is exposed, which has to be considered as a further class in the classification step.

Available orthophoto time series from public surveys help to reconstruct the development of the landslides. The initial trigger can be dated to the year 1965, when heavy rainfalls caused the landslides to move for the first time. Today the areas are prone to ongoing erosion processes and reactivation of landsliding, which precludes a reestablishment of closed vegetation by natural succession. In order to better understand these bio-geomorphological dynamics, terrestrial laser scans (TLS) were conducted between 2011 and 2018 on a biannual basis (Spring and Autumn). During this period the first five measurements were conducted with an OPTECH Inc. Iiris 3D instrument and the other measurements with a RIEGL VZ6000 instrument. Both scanners have different characteristics in terms of wavelength, recorded intensity, and scanning geometry. Thus, even after radiometric correction, the LiDAR intensity data cannot be compared straightforwardly between all epochs.

Andreas Mayr and Clemens Geitner are with the University of Innsbruck, Institute of Geography, Innrain 52, 6020 Innsbruck, Austria (andreas.mayr@uibk.ac.at, clemens.geitner@uibk.ac.at).

Martin Rutzinger is with the Austrian Academy of Sciences, Institute for Interdisciplinary Mountain Research, Technikerstr. 21a, 6020 Innsbruck, Austria (martin.rutzinger@oeaw.ac.at).

Photogrammetric Engineering & Remote Sensing
Vol. 85, No. 6, June 2019, pp. 455–462.
0099-1112/19/455–462

© 2019 American Society for Photogrammetry
and Remote Sensing
doi: 10.14358/PERS.85.6.455

Georeferencing of the TLS point clouds is performed by Differential Global Navigation Satellite System (DGNS) measured targets, but, depending on the agricultural management cycle (and thus the height of the grass), access to the site was at times restricted. This precluded the use of GNSS surveyed temporary targets for some epochs. In addition, the site offers only limited possibilities to install targets permanently as bedrock outcrops exist only in the lower part of the site. Therefore, the coregistration of the point cloud time series relies largely on matching with iterative closest point adjustment (Besl and McKay 1992). For each epoch, point clouds were acquired from two predefined scan positions.

These point cloud pairs are coregistered by matching specific patches of the point clouds, which have been defined to cover variable surface orientations and spread across the scene. Then, point clouds from different epochs are coregistered by matching surface patches outside the landslide areas. All point clouds are thinned out, and thus homogenized in terms of point spacing, by 3D block filtering.

Workflow and Methods

This work is based on two main modules (Figure 2), (i) an automated workflow for landslide-related object classification

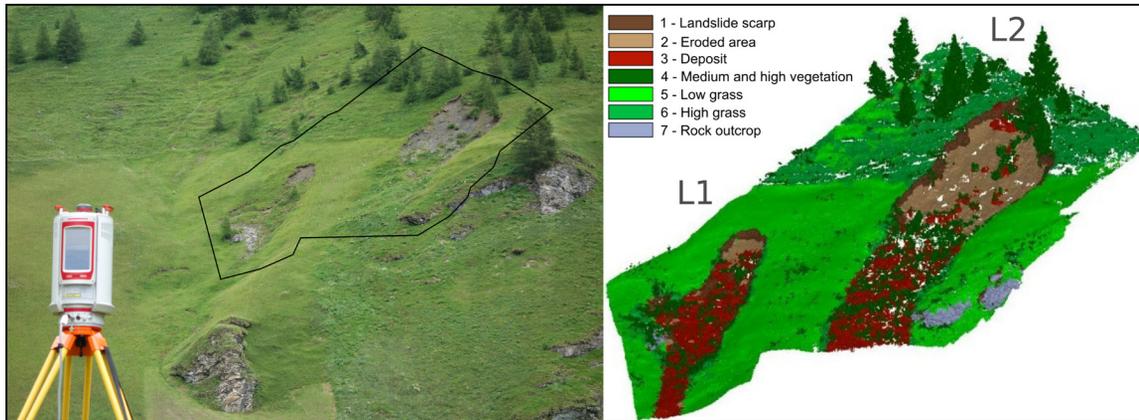


Figure 1. Landslides at the test site (left) and classified point cloud showing landslide 1 (L1) and landslide 2 (L2) (right).

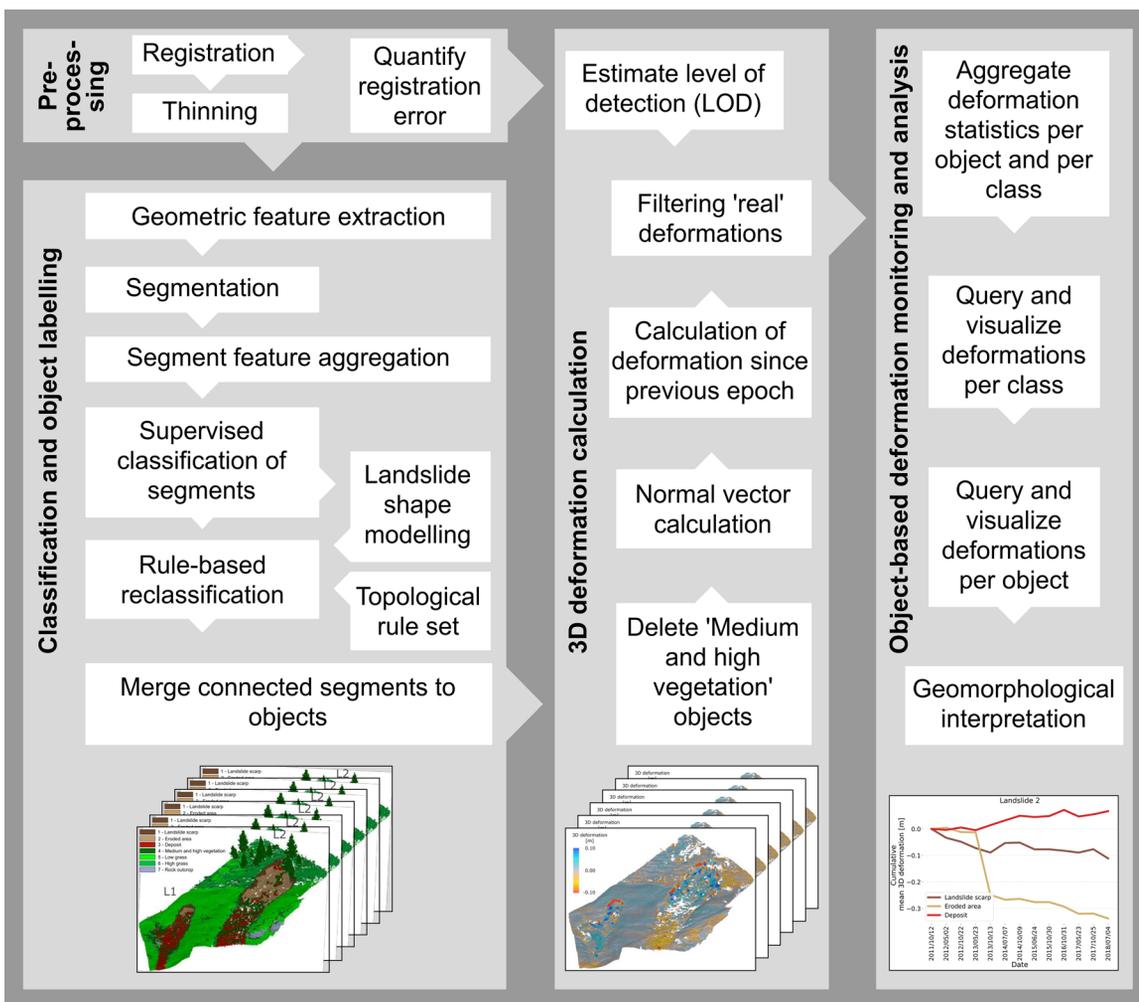


Figure 2. Schematic overview of the object-based landslide monitoring approach.

in 3D point clouds (Mayr *et al.* 2017), and (ii) 3D deformation calculations (Lague *et al.* 2013; Fey and Wichmann 2017). These two modules are now combined in a four-dimensional object-based approach for multi-temporal identification and analysis of landslide changes, using the complete series of TLS epochs that have been acquired at the test site so far. Python scripts, together with the free open-source software System for Automated Geoscientific Analysis (SAGA) (Conrad *et al.* 2015) and the Laserdata Information System (LIS) tools extension for SAGA (Rieg *et al.* 2014), are used to implement the methods for classification and change analysis of point cloud objects. These methods are described in the following sections.

Point Cloud Classification

The applied classification approach, developed by Mayr *et al.* (2017), works directly with the 3D point cloud and the features derived from it. Intensity data has not been considered because of the different characteristics of the two scanning systems used. In a first step, geometric point cloud features are derived. These are two-dimensional (2D) z-range, 3D/2D point density ratio, slope, slope distance (SD) from plane, omnivariance, and curvature. These features are computed for three neighborhood radii, 0.2, 0.4, and 1.0 m. Based on the geometric features, the points are grouped into homogeneous segments by seeded region growing. The result is an oversegmented point cloud (partitioned into subobjects), which is the input for the subsequent classification step, when these subobject segments are assigned to a target classification scheme. This scheme has seven classes: “scarp”, “eroded area”, “deposit”, “rock outcrop”, “low grass”, “high grass”, “medium and high vegetation”.

The classification itself is a two-step process. First, a supervised classification is performed by building a random forest classifier (Breiman 2001) from the geometric features described above. For this purpose, the target classes are manually labelled in one reference point cloud, where a subset with L2 and its surroundings serves as training and a subset containing L1 as validation point cloud. The trained classifier then labels all other 12 epochs based on the geometric features of their segments. In a second step, the classification result is enhanced by deriving landslide outlines. Classified scarp segments are filtered by size and connectivity to remove outliers and define the top of the landslides. Assuming that gravity forced the mass movement downslope, the process area is laterally limited by flow routing on the gridded TLS terrain model. These upper and lateral landslide outlines are used for applying topological rules to correct for classes that, per definition, cannot exist inside or outside the landslide. The entire classification workflow, including the machine-learning part and the topological rules, is described fully in Mayr *et al.* (2017).

Point Cloud Deformation Calculation

Prior to the deformation calculation, medium and high vegetation is deleted from the classified point clouds to keep only points which are relevant for geomorphological change analysis. In the next step, a random sample consensus-approach (Fischler and Bolles 1981) fits planes robustly to points of a spherical neighborhood around each point and derives this point’s normal vector. Then, deformations of a point cloud epoch, compared to the previous epoch, are estimated by calculating the distance between these two point clouds. This is done with the Fey and Wichmann (2017) implementation of the Multiscale Model to Model Cloud Comparison (Lague *et al.* 2013). The required parameters for the test site are determined by empirical testing. For every interest point in point cloud *A*, the corresponding closest point along the normal is searched in point cloud *B*. The search is restricted by a cylindrical neighborhood definition around the normal vector

(radius = 0.2 m and length = 2.0 m). For the matched point pairs from the two epochs, point sets are selected in a neighborhood of 0.5 m radius. Planes are fitted to these point sets by minimizing the squares of the residuals for all points that deviate by less than 45° from the interest point’s normal vector.

To measure the 3D distance, the two interest points are projected on their corresponding planes and, starting from the projected point from epoch *A*, the distance is measured, either along this point’s normal vector or in the opposite direction, up to the intersection with the local plane in epoch *B*. Depending on the required direction for a valid measurement, the resulting distance is accompanied by a negative or positive sign. This indicates if the surface of epoch *A* is below, inside or behind the surface of epoch *B* (negative sign) or if the surface of epoch *A* is above, outside or in front of the surface of epoch *B* (positive sign).

Object Changes

In the subsequent object modelling step, adjacent segments are merged to one geomorphological object of interest if they have an identical class label. This step also considers the previously estimated distances, compared to the previous epoch as a criterion to separate stable from unstable (i.e., changing) landslide parts. For subsequent analyses, a straight (manually defined) 2D boundary separates landslides L1 and L2, including their respective surroundings.

For each epoch, the average 3D deformation is aggregated per class and per object, using the attributes linked to the single laser points (i.e., the 3D distance, the class label, and a unique object identifier). Here, only 3D distances approved as “real” changes are considered, that is, with a magnitude exceeding the level-of-detection (in the section “Level of Detection and Parameter Test”). The class-based and object-based mean deformations are stored as attributes of the point clouds. Finally, both the class-based and the object-based information from all epochs is aggregated in one data frame for each landslide. This is the basis for all further time series analysis and, together with the point clouds and their attributes, for the geomorphological interpretation of changes.

Level of Detection and Parameter Test

As a spatially explicit estimate for the reliability of detected changes, allowing for the influence of data noise and systematic deviations due to registration errors, the method used for distance calculation (Lague *et al.* 2013; Fey and Wichmann 2017) derives the level-of-detection at the 95% confidence interval (LOD_{95}). Used as a threshold to consider only “real” changes for further analysis and interpretation (magnitude of 3D distances > LOD_{95}), this level-of-detection is defined as

$$LOD_{95} = \pm 1.96 \sqrt{\left(\frac{\sigma_A^2}{n_A}\right) + \left(\frac{\sigma_B^2}{n_B}\right)} + reg \quad (1)$$

where *A*, *B* = point cloud epochs,
 σ^2 = plane fitting variance,
n = number of points in the fitting neighborhood,
reg = registration error.

In addition to the number of points (*n*) in the point set and the variance of their distances to the fitted plane (σ^2), the LOD_{95} is determined by the registration error *reg*. For the presented case study, we quantify the registration error between two subsequent epochs by calculating their distance at surface patches, which are distributed over the test site and assumed to be stable. The absolute mean distance plus standard deviation of these distance measurements is used as registration error in the LOD_{95} calculation (Table 1).

Table 1. Distance measurements for each epoch relative to the previous epoch, calculated at the accuracy control surface patches. The absolute mean distance plus standard deviation is used as an estimate for the registration error in the 3D deformation calculation step.

PC	Distance mean [m]	Distance sd [m]	Registration error <i>reg</i>
			Distance mean + sd [m]
1	-	-	-
2	0.003	0.018	0.020
3	-0.007	0.023	0.030
4	-0.001	0.025	0.025
5	0.008	0.024	0.032
6	0.009	0.066	0.075
7	-0.002	0.043	0.046
8	-0.013	0.043	0.056
9	0.024	0.037	0.060
10	0.003	0.024	0.027
11	-0.015	0.032	0.047
12	0.017	0.037	0.054
13	0.002	0.049	0.050
all [mean]	0.002	0.035	0.044

To illustrate the sensitivity of the class-based change analysis to this parameter *reg*, distances for the first nine epochs are computed with registration errors (*reg*) between 0.02 m to 0.16 m, increased incrementally by 0.02 m. The influence of these estimated error values is further analyzed for the class “landslide scarp” (see the section “Level of Detection—Parameter Sensitivity and Impact on the Object-Based Change”).

Experimental Results and Discussion

Object Changes and Interpretation

The mean deformations per object and per class are described by 3D distances between point cloud epochs, accompanied by a sign (+/-) for the direction of surface deformation. Figure 3 summarizes these changes for the entire monitoring period, showing higher deformations for the landslide-related objects (“landslide scarp”, “eroded area”, and “deposit”) than for “low grass” and “rock outcrop”. Deformed “rock outcrop” objects must be attributed to either registration errors ($> LOD_{95}$) or to classification errors. The latter can occur due to similarities of small rock cliffs with “scarp” areas and “medium and high vegetation” in a geometric feature space (e.g. with relation to slope gradient or 2D z-range (Mayr *et al.* 2017)). Decimeter-scale changes of the “high grass” class can mostly be attributed to the growing cycle and occasional mowing. Less prevalent large changes of this class, however, point to false positives in the classification, such as misclassified segments belonging to trees and shrubs. Focusing on landslide-related objects, landslide L2 changed more than landslide L1 during the monitoring period, with 50% of the L2 “eroded area” objects changing by more than 0.5 m (surface retreat). In landslide L1, only the “scarp” objects returned substantial deformations. In both landslides, “scarp” objects not only show negative deformations, as an indicator for scarp retreat by erosion, but also positive deformations in the “scarp” class were measured. The latter can be interpreted either as collapsed and dislocated (former) “scarp” parts (now correctly “deposit”) being misclassified as “scarp”, or as scarp parts that moved only slightly and reside in the scarp area (i.e., upper landslide boundary) before collapse and further dislocation by gravity. The depositional zone should (in theory) be characterized by prevailing positive surface deformation due to mass gain by deposition. However, the “deposit” class also shows negative deformations (i.e., surface retreat or lowering). This indicates that either settlement in unconsolidated deposits occurs locally or that objects in the depositional zone may continue to be transported further downslope. Field observations suggest that the erosion, transport, and deposition processes become more diffuse over time, with turf and soil clods disintegrating and material being washed out by surface runoff and transported out of the test site.

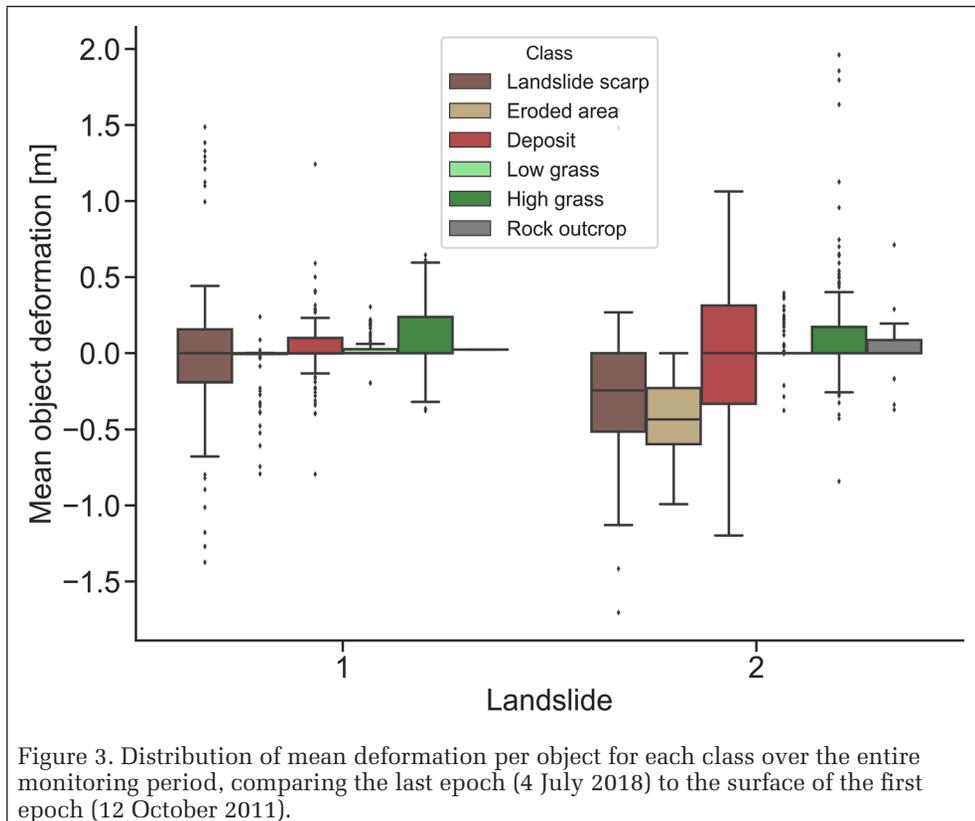


Figure 3. Distribution of mean deformation per object for each class over the entire monitoring period, comparing the last epoch (4 July 2018) to the surface of the first epoch (12 October 2011).

Figure 4 shows the development of each landslide for the individual epochs of the time series, with mean deformations per landslide-related class. The most substantial change during the monitoring period is a reactivation of landslide L2 in summer 2013. Between 23 May 2013 and 13 October 2013, either superficial erosion or subsidence lowered the surface in the upper part (negative deformations), while material accumulated in the depositional area (positive deformations). This accelerated movement was most likely triggered by heavy rain in May and June 2013, as indicated by precipitation measurements of 110 mm within three days at a nearby weather station (Figure 5). This event contrasts with a less dynamic development of landslide L2 throughout the remaining

Figure 4 shows the development of each landslide for the individual epochs of the time series, with mean deformations per landslide-related class. The most substantial change during the monitoring period is a reactivation of landslide L2 in summer 2013. Between 23 May 2013 and 13 October 2013, either superficial erosion or subsidence lowered the surface in the upper part (negative deformations), while material accumulated in the depositional area (positive deformations). This accelerated movement was most likely triggered by heavy rain in May and June 2013, as indicated by precipitation measurements of 110 mm within three days at a nearby weather station (Figure 5). This event contrasts with a less dynamic development of landslide L2 throughout the remaining

period. Future work could integrate semantically differentiated landslide monitoring information and precipitation time series more thoroughly to relate detailed landslide dynamics to hydro-meteorological conditions.

At both landslides, retrogressive erosion at the scarp is represented by a negative trend of deformation (Figure 4). At the scarp, the soil below the root zone has been gradually eroded (probably by water), causing parts of the overhanging scarp to collapse occasionally and slide or topple downwards. Figure 6 shows an example from landslide L1, where “landslide scarp” objects broke off and were deposited several meters below.

On a multiple-year to decadal timescale, these small recurrent events cause the eroded areas to expand in an uphill direction, a phenomenon which has also been observed in aerial orthophoto time series. The collapsed parts of the scarp are deposited (temporarily) in the landslides’ eroded area or depositional zone, which is indicated by positive deformations in these areas (Figures 3, 4, and 5). In total, however, the deposition of material is underrepresented by the object- and class-based deformation analysis. As explained before, this is attributed to (i) continued transport and (ii) dispersion of material, with absolute surface deformations being lower than the LOD_{95} . At least in the eroded areas, the continuing or recurrent processes of erosion, transport, and deposition delay vegetation succession and with it soil formation.

Level of Detection—Parameter Sensitivity and Impact on the Object-Based Change

The epoch-based registration error (i.e., between the two scans of one epoch) was assessed based on a distance calculation at independent surface patches (i.e., not used for registration). It tends to be lower than the multi-epoch registration error, probably because of consistent vegetation height and density. The epoch-based registration error is included in the spatially variable level-of-detection via the local plane fitting variance (σ^2). In overlapping areas, registration errors result in a higher plane fitting variance and, hence, in a higher level-of-detection.

The registration accuracy between subsequent epochs of the presented test data, estimated from surface patches outside the landslides, varies from 0.020 m to 0.075 m between the epochs (Table 1). A lack of well distributed smooth areas that are stable and unvegetated constrains the possibilities for an accurate registration of all point cloud epochs. In addition to uncertainties in the semantic classification of point cloud objects, this limits the reliability of the presented results. The applied deformation calculation method (Lague *et al.* 2013; Fey and Wichmann 2017) accounts for this by excluding deformations smaller than the level-of-detection at the 95% confidence interval (LOD_{95}). To demonstrate how this affects the results, the sensitivity of class-based deformations to the registration error reg was tested using a range of reg values for the deformation calculation.

The LOD_{95} increases with higher registration errors (Equation 1), (Table 2). Consequently, more (per-point) deformations fall below the higher local LOD_{95} and are excluded from

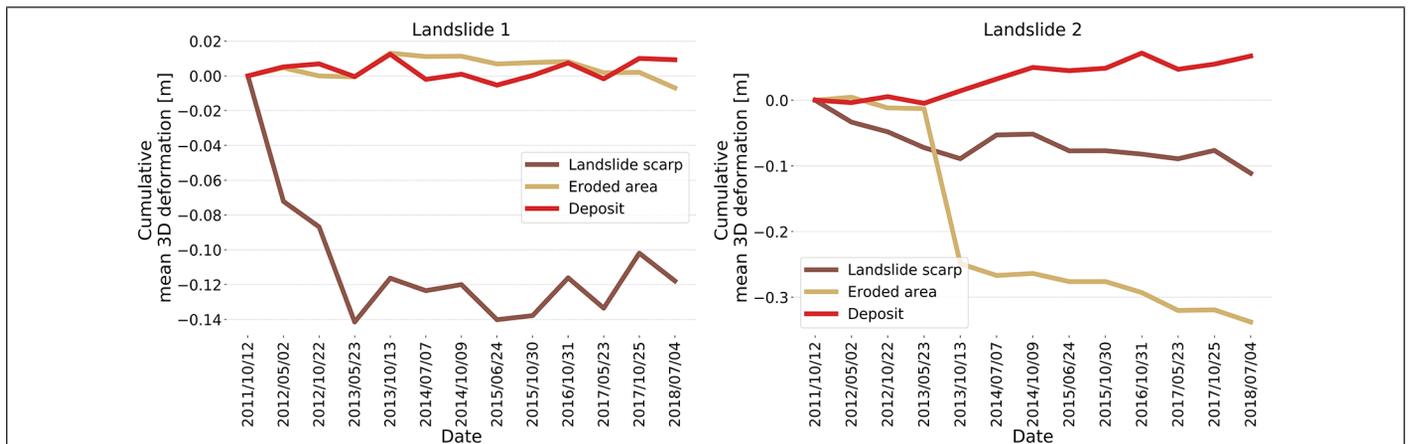


Figure 4. Cumulative mean 3D deformation per class for each of the two landslides. Note the difference in y-axis scale.

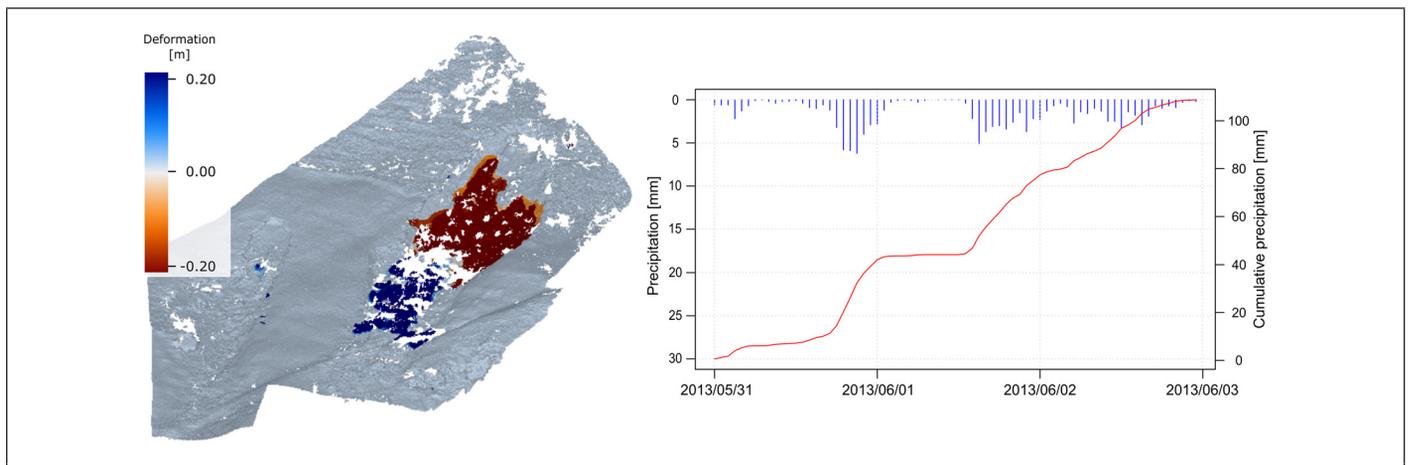


Figure 5. Oblique view of the epoch 13 October 2013, colored by mean 3D deformation per object compared to 23 May 2013 (left). While the left landslide L1 remained mostly stable, the sliding mass of landslide L2 was reactivated. An intense precipitation event probably triggered the reactivation. Hourly precipitation and cumulative precipitation over three days in early summer 2013 were recorded by a rain gauge close to the test site (right; data provided by Hydrographic Service of Tyrol).

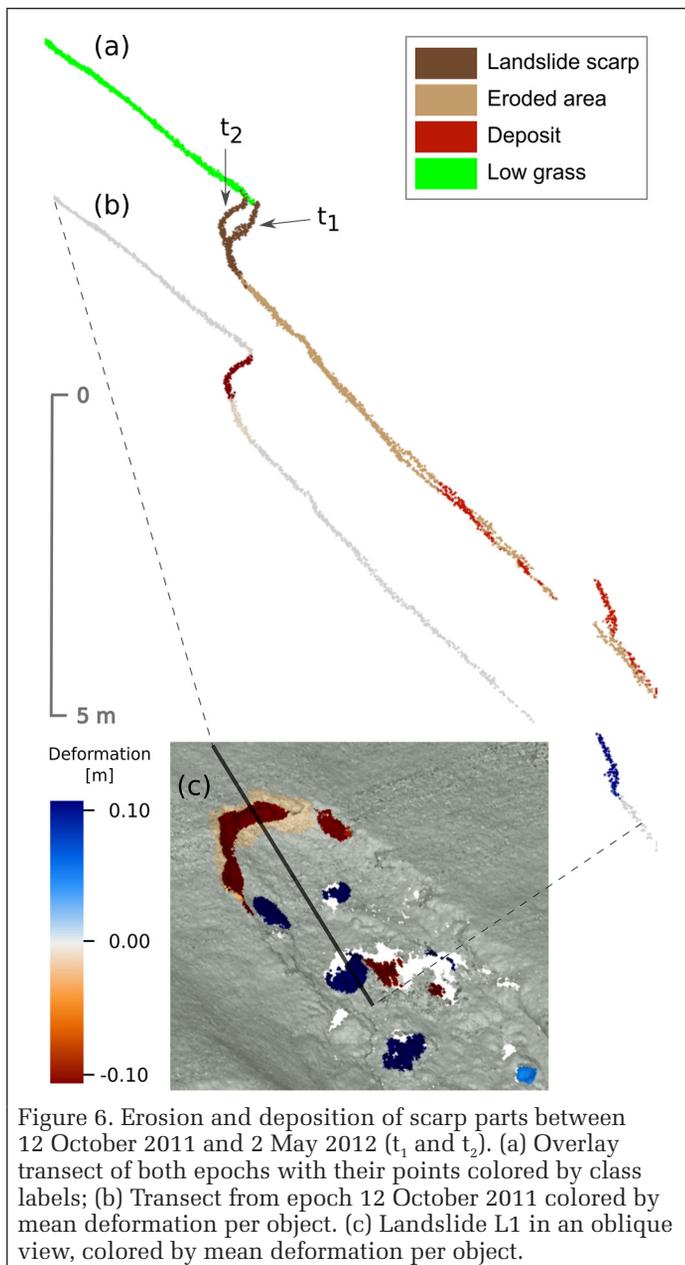


Figure 6. Erosion and deposition of scarp parts between 12 October 2011 and 2 May 2012 (t_1 and t_2). (a) Overlay transect of both epochs with their points colored by class labels; (b) Transect from epoch 12 October 2011 colored by mean deformation per object. (c) Landslide L1 in an oblique view, colored by mean deformation per object.

an object- and class-based aggregation. For example, in epoch 8 a registration error $reg = 0.10$ m resulted in a mean LOD_{95} of 0.200 m (with a standard deviation of 0.002 m), compared to a mean LOD_{95} of 0.082 m for $reg = 0.04$ m. Assuming a constant registration error, the LOD_{95} is relatively homogeneous (low standard deviations) between epochs as well as within each epoch (Table 2). Smaller registration errors decrease the LOD_{95} as a threshold for “real” change, and in turn more deformations are included for the object- and class-based calculation of mean deformation. This results in larger mean deformation magnitudes for some epochs, as demonstrated by the test results shown in Figure 7.

If an object or class includes points without deformation or deformation below the LOD_{95} , the averaging per object or class can result in a mean object deformation magnitude considerably below the LOD_{95} . To some degree, such effects are mitigated by including the deformation as a constraint in the object modelling step (see sections “Point Cloud Classification” and “Point Cloud Deformation Calculation”). Certain epochs show systematic variations of object- and class-based

Table 2. LOD_{95} statistics per point cloud epoch for two different registration errors (reg).

	reg [m]			
	0.04		0.10	
Epoch	2	8	9	8
Date	2 May 2012	24 June 2015	30 October 2015	2 June 2015
LOD_{95} [m]	min	0.079	0.079	0.079
	max	0.123	0.141	0.141
	mean	0.081	0.082	0.081
	sd	0.002	0.002	0.001

deformations if these are computed with different reg values, which determine the LOD_{95} and, accordingly, define “real” change. For higher values of reg , only a generalized trend line for the calculated deformations per class is obtained. In contrast, analyses based on smaller registration errors (i.e., with a less aggressive filtering of deformations) suggest more diverse process dynamics and higher magnitudes of change.

Discussion

The registration sensitivity test underlines the relevance of a registration accuracy assessment to effectively exclude systematic (registration) errors by defining a realistic level of detection. Unfortunately, the conditions for an accurate registration and its assessment are often poor in natural environments, because stable areas with smooth and differently oriented surfaces are scarce. In our test data set, multi-temporal point cloud registration and its error assessment were based on surface patches outside the landslide process zone. These patches contained (i) small rock walls and, to improve the spatial distribution, (ii) parts of the meadow where no signs of geomorphological activity (on a relevant scale) were visible. In the meadow patches, however, the spatially and temporally variable grass height compromised the registration accuracy and the assessment of errors. This variability is related to (i) seasonal changes of the vegetation and (ii) grassland management (mowing). Since the changes in grass height between epochs are not always consistent across the test site, some patches can show comparatively large distances, which are not caused by a registration error. This impacts on the distributions of derived registration errors, as found by the patch-based distance analysis, which deviate from a Gaussian distribution to some degree. Therefore, assuming a Gaussian distribution for modelling registration errors may not ideally reflect the real situation, but we consider it still an adequate (and viable) method to get an estimate of the registration error. One strategy for dealing with the uncertainty of this registration error surrogate could be to additionally perform the class-based deformation analysis with a range of more conservative registration errors. This might help to get an estimate of the uncertainty of the results to be considered for geomorphological interpretations.

The second major source for uncertainties in the proposed approach is the classification of objects. More consistent input point clouds, for instance captured with comparable sensors and acquisition settings (regarding exact scan positions, scan resolution, LiDAR wavelength, etc.) in a time series, would partly mitigate this. Integration of radiometric features (LiDAR reflectance and spectral values from images) or LiDAR pulse shape features for the classification step might address the problem of geometric variability and ambiguity inherent in many natural objects. In addition, calculated deformations could be considered in the classification to obtain class definitions that are more process-oriented and do not only rely on the geometry of objects and their topology. The performance

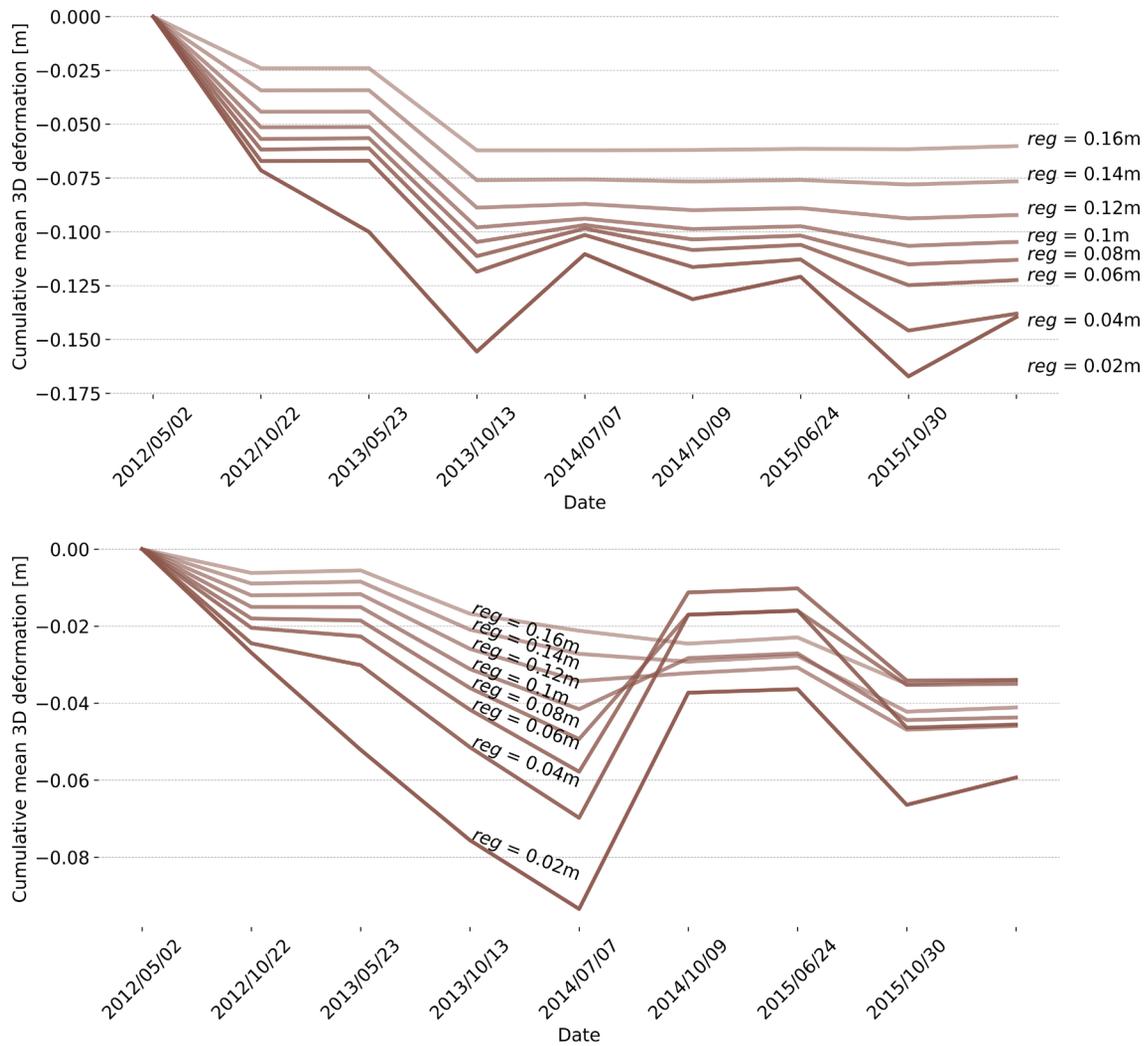


Figure 7. Cumulative “landslide scarp” mean deformation for landslides L1 (top) and L2 (bottom), calculated with different registration errors (*reg*).

and accuracy of the classification approach is addressed in Mayr *et al.* (2017) which is, however, based on an assessment of one epoch only. A quantitative accuracy assessment of classification errors across the entire time series is unfortunately not possible at the moment due to a lack of multi-temporal validation data. Future studies could address the challenge of combining semantic classification errors and errors from a deformation calculation into a total error budget for object-based 3D monitoring.

Despite these problems, the experimental results illustrate the potential of an object-based approach for automated geomorphological information extraction from point cloud time series. On the one hand, the erosion or deposition of individual objects of a landslide system can be detected and analyzed (such as detached parts of the scarp; Figure 6). On the other hand, more general development trends for distinct object classes, such as “landslide scarp”, “eroded area”, or “deposit”, can be aggregated from the complex point cloud time series (Figures 3 and 4).

The focus of this paper is on the detection of surface deformations and the semantic interpretation, rather than their quantification. The aggregated average deformation per class or object cannot be directly interpreted quantitatively, but rather as an indicator of change. The automated

data processing pipeline and visualizations of the aggregated results can help to get an overview and identify the timing and location of the most prominent events (Figure 4) from a large multi-temporal data set. For the respective time steps, the object classes of interest can then be analyzed in detail, using the 3D point cloud representation with labelled object changes (Figure 6). A volumetric quantification of the eroded or deposited material can be performed (i) in 2.5D by raster-based DTM differencing (e.g. Stumpf *et al.* 2015), or (ii) by constructing 3D alpha shapes between points of two epochs for a certain class. For the latter, an example of landslide scarp erosion is given in Mayr *et al.* (2017). Finally, automated object tracking and trajectory reconstruction might characterize larger dislocations of objects more appropriately than surface distances along surface normals or volumetric change analyses.

Conclusions

This paper presents an approach to monitoring landslides at the level of geomorphological objects by automated point cloud analysis. We demonstrate how information on geomorphological processes can be extracted from the point cloud time series and aggregated at more abstract levels. A case

study revealed the diversity of erosion, transport, and deposition processes involved in secondary landslide development (i.e., after initial slope failure). Point cloud registration errors and errors of a geometry-based classification are rated as the two main limitations for object-based monitoring of complex natural scenes. However, the results of our analysis identify retrogressive erosion of the scarp and episodic downslope movement of the sliding mass as major processes of secondary landslide development. This largely agrees with direct field observations at the investigated slope, but also with conceptual landslide models. Thus, we conclude that multi-temporal object-based point cloud analysis is a promising approach to (i) reduce the spatio-temporal complexity of many environmental processes and (ii) utilize the details and accuracy provided by laser scanning. Landslide investigators, who have to interpret deformation measurements in terms of geomorphological changes, can benefit from such tools in future. Joint analysis of semantically interpreted spatio-temporal information on slope deformations with their potential drivers (such as precipitation and soil structure) will improve the understanding and assessment of erosion and landslide processes.

Acknowledgments

This work was carried out within (i) the project ERODYN (Shallow erosion dynamics in mountain grasslands of South Tyrol: Monitoring, process analysis and mitigation measures, <http://www.mountainresearch.at/erodyn/>), funded by the Autonomous Province of Bozen—South Tyrol (Italy), and (ii) the Doctoral Fellowship Program of the Austrian Academy of Sciences. We thank all colleagues who supported the data acquisition campaigns in the field.

References

- Alewell, C., M. Egli and K. Meusburger. 2015. An attempt to estimate tolerable soil erosion rates by matching soil formation with denudation in Alpine grasslands. *Journal of Soils and Sediments* 15(6):1383–1399.
- Besl, P. J. and N. D. McKay. 1992. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14(2):239–256.
- Breiman, L. 2001. Random forests. *Machine Learning* 45(1):5–32.
- Conrad, O., B. Bechtel, M. Bock, H. Dietrich, E. Fischer, L. Gerlitz, J. Wehberg, V. Wichmann and J. Böhner. 2015. System for automated geoscientific analyses (SAGA) v. 2.1.4. *Geoscientific Model Development* 8(7):1991–2007.
- Cramer, M., N. Haala, D. Laupheimer, G. Mandlbürger and P. Havel. 2018. Ultra-high precision UAV-based LiDAR and dense image matching. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-1*:115–120.
- Fey, C. and V. Wichmann. 2017. Long-range terrestrial laser scanning for geomorphological change detection in Alpine terrain—Handling uncertainties. *Earth Surface Processes and Landforms* 42:789–802.
- Fischler, M. A. and R. C. Bolles. 1981. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM* 24(6):381–395.
- Gira, P., N. Pfeifer and G. Mandlbürger. 2016. Rigorous strip adjustment of UAV-based laserscanning data including time-dependent correction of trajectory errors. *Photogrammetric Engineering and Remote Sensing* 82(12):945–954.
- Jaboyedoff, M., T. Oppikofer, A. Abellán, M. H. Derron, A. Loye, R. Metzger and A. Pedrazzini. 2012. Use of LIDAR in landslide investigations: A review. *Natural Hazards* 61(1):5–28.
- Kjekstad, O. and L. Highland. 2009. Economic and social impacts of landslides. In *Landslides—Disaster Risk Reduction*, edited by K. Sassa and P. Canuti, 573–587.
- Kromer, R. A., A. Abellán, D. J. Hutchinson, M. Lato, M. A. Chanut, L. Dubois and M. Jaboyedoff. 2017. Automated terrestrial laser scanning with near-real-time change detection—Monitoring of the Séchilienne landslide. *Earth Surface Dynamics* 5(2):293–310.
- Lague, D., N. Brodu and J. Leroux. 2013. Accurate 3D comparison of complex topography with terrestrial laser scanner: Application to the Rangitikei canyon (NZ). *ISPRS Journal of Photogrammetry and Remote Sensing* 82:10–26.
- Lucieer, A., S. M. D. Jong and D. Turner. 2014. Mapping landslide displacements using Structure from Motion (SfM) and image correlation of multi-temporal UAV photography. *Progress in Physical Geography* 38(1):97–116.
- Mayr, A., M. Rutzinger, M. Bremer, S. Oude Elberink, F. Stumpf and C. Geitner. 2017. Object-based classification of terrestrial laser scanning point clouds for landslide monitoring. *The Photogrammetric Record* 32(160):377–397.
- Mayr, A., M. Rutzinger and C. Geitner. 2018. Multitemporal analysis of objects in 3D point clouds for landslide monitoring. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-2*, 42(2):691–697.
- Petley, D. 2012. Global patterns of loss of life from landslides. *Geology* 40(10):927–930.
- Rieg, L., V. Wichmann, M. Rutzinger, R. Sailer, T. Geist and J. Stötter. 2014. Data infrastructure for multitemporal airborne LiDAR point cloud analysis—Examples from physical geography in high mountain environments. *Computers, Environment and Urban Systems* 45:137–146.
- Scaioni, M., L. Longoni, V. Melillo and M. Papini. 2014. Remote sensing for landslide investigations: An overview of recent achievements and perspectives. *Remote Sensing* 6(10):9600–9652.
- Scaioni, M., T. Feng, P. Lu, G. Qiao, X. Tong, R. Li, L. Barazzetti, M. Previtali and R. Roncella. 2015. Close-range photogrammetric techniques for deformation measurement: Applications to landslides. In *Modern Technologies for Landslide Monitoring and Prediction*, edited by M. Scaioni, 13–41.
- Stumpf, A., J. P. Malet, P. Allemand, M. Pierrot-Deseilligny and G. Skupinski. 2015. Ground-based multi-view photogrammetry for the monitoring of landslide deformation and erosion. *Geomorphology* 231:130–145.
- Turner, A. K. and R. L. Schuster. 1996. Landslides: Investigation and mitigation. Transportation Research Board Special Report 247. National Research Council, Washington, D.C.
- Vericat, D., M. W. Smith and J. Brasington. 2014. Patterns of topographic change in sub-humid badlands determined by high resolution multi-temporal topographic surveys. *Catena* 120:164–176.
- Wiegand, C. and C. Geitner. 2010. Shallow erosion in grassland areas in the Alps. What we know and what we need to investigate further. In *Challenges for Mountain Regions—Tackling Complexity*, edited by A. Borsdorf, G. Grabherr, K. Heinrich, B. Scott, and J. Stötter, 76–83.

Advertise in the #1 publication in the imaging and geospatial information industry!



Impact Factor (2018): 3.15

PE&RS (Photogrammetric Engineering & Remote Sensing) is the official journal of the American Society for Photogrammetry and Remote Sensing—the Imaging and Geospatial Information Society (ASPRS). This highly respected publication is the #1 publication in the industry by a wide margin. Advertisers with a desire to reach this market can do no better than PE&RS. It delivers the right audience of professionals who are loyal and engaged readers, and who have the purchasing power to do business with you.

PE&RS Readership Highlights

Circulation: 3,000

Total audience: 6,400*

Founded in 1934, the American Society for Photogrammetry and Remote Sensing (ASPRS) is a scientific association serving professional members throughout the world. Our mission is to advance knowledge and improve understanding of mapping sciences to promote the responsible applications of photogrammetry, remote sensing, geographic information systems (GIS), and supporting technologies.

Our members are analysts/specialists, educators, engineers, managers/administrators, manufacturers/ product developers, operators, technicians, trainees, marketers, and scientists/researchers. Employed in the disciplines of the mapping sciences, they work in the fields of Agriculture/Soils, Archeology, Biology, Cartography, Ecology, Environment, Forestry/Range, Geodesy, Geography, Geology, Hydrology/Water Resources, Land Appraisal/ Real Estate, Medicine, Transportation, and Urban Planning/Development.

RESERVE YOUR SPACE TODAY!

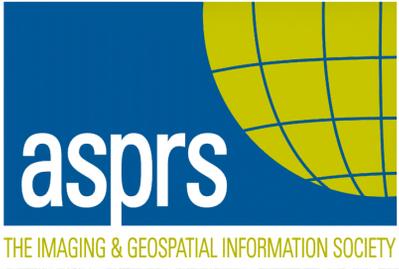
Bill Spilman, ASPRS Advertising, Exhibit Sales & Sponsorships
 (877) 878-3260 toll-free
 (309) 483-6467 direct
 (309) 483-2371 fax
 bill@innovativemediasolutions.com

	Corporate Member Exhibiting at a 2019 ASPRS Conference	Corporate Member	Exhibitor	Non Member
<i>All rates below are for four-color advertisements</i>				
Cover 1	\$1,850	\$2,000	\$2,350	\$2,500
<i>In addition to the cover image, the cover sponsor receives a half-page area to include a description of the cover (maximum 500 words). The cover sponsor also has the opportunity to write a highlight article for the journal. Highlight articles are scientific articles designed to appeal to a broad audience and are subject to editorial review before publishing. The cover sponsor fee includes 50 copies of the journal for distribution to their clients. More copies can be ordered at cost.</i>				
Cover 2	\$1,500	\$1,850	\$2,000	\$2,350
Cover 3	\$1,500	\$1,850	\$2,000	\$2,350
Cover 4	\$1,850	\$2,000	\$2,350	\$2,500
Advertorial	1 Complimentary Per Year	1 Complimentary Per Year	\$2,150	\$2,500
Full Page	\$1,000	\$1,175	\$2,000	\$2,350
2 page spread	\$1,500	\$1,800	\$3,200	\$3,600
2/3 Page	\$1,100	\$1,160	\$1,450	\$1,450
1/2 Page	\$900	\$960	\$1,200	\$1,200
1/3 Page	\$800	\$800	\$1,000	\$1,000
1/4 Page	\$600	\$600	\$750	\$750
1/6 Page	\$400	\$400	\$500	\$500
1/8 Page	\$200	\$200	\$250	\$250
Other Advertising Opportunities				
Wednesday Member Newsletter Email Blast	1 Complimentary Per Year	1 Complimentary Per Year	\$600	\$600

A 15% commission is allowed to recognized advertising agencies

THE MORE YOU ADVERTISE THE MORE YOU SAVE! PE&RS offers frequency discounts. Invest in a three-times per year advertising package and receive a 5% discount, six-times per year and receive a 10% discount, 12-times per year and receive a 15% discount off the cost of the package.

*Based on 2 readers per copy as well as online views | Source: PE&RS Readership Survey, Fall 2012



asprs

THE IMAGING & GEOSPATIAL INFORMATION SOCIETY

ASPRS AERIAL DATA CATALOG

“THE SOURCE FOR FINDING AERIAL COLLECTIONS”

[HTTP://DPAC.ASPRS.ORG](http://dpac.asprs.org)

The ASPRS Aerial Data Catalog is a tool allowing owners of aerial photography to list details and contact information about individual collections.

By providing this free and open metadata catalog with no commercial interests, the Data Preservation and Archiving Committee (DPAC) aims to provide a definitive metadata resource for all users in the geospatial community to locate previously unknown imagery.

DPAC hopes this Catalog will contribute to the protection and preservation of aerial photography around the world!

ASPRS Members: We Need Your Help!
There are three ways to get involved

1

USE

Use the catalog to browse over 5,000 entries from all 50 states and many countries. Millions of frames from as early as 1924!

2

SUPPLY

Caretakers of collections, with or without metadata, should contact DPAC to add their datasets to the catalog free of charge!

3

TELL

Spread the word about the catalog! New users and data collections are key to making this a useful tool for the community!

For More Details or To Get Involved Contact:

DAVID RUIZ • DRUIZ@QUANTUMSPATIAL.COM • 510-834-2001 OR DAVID DAY • DDAY@KASURVEYS.COM • 215-677-3119

LEARN
DO
GIVE
BELONG

ASPRS Offers

- » Cutting-edge conference programs
- » Professional development workshops
- » Accredited professional certifications
- » Scholarships and awards
- » Career advancing mentoring programs
- » *PE&RS*, the scientific journal of ASPRS

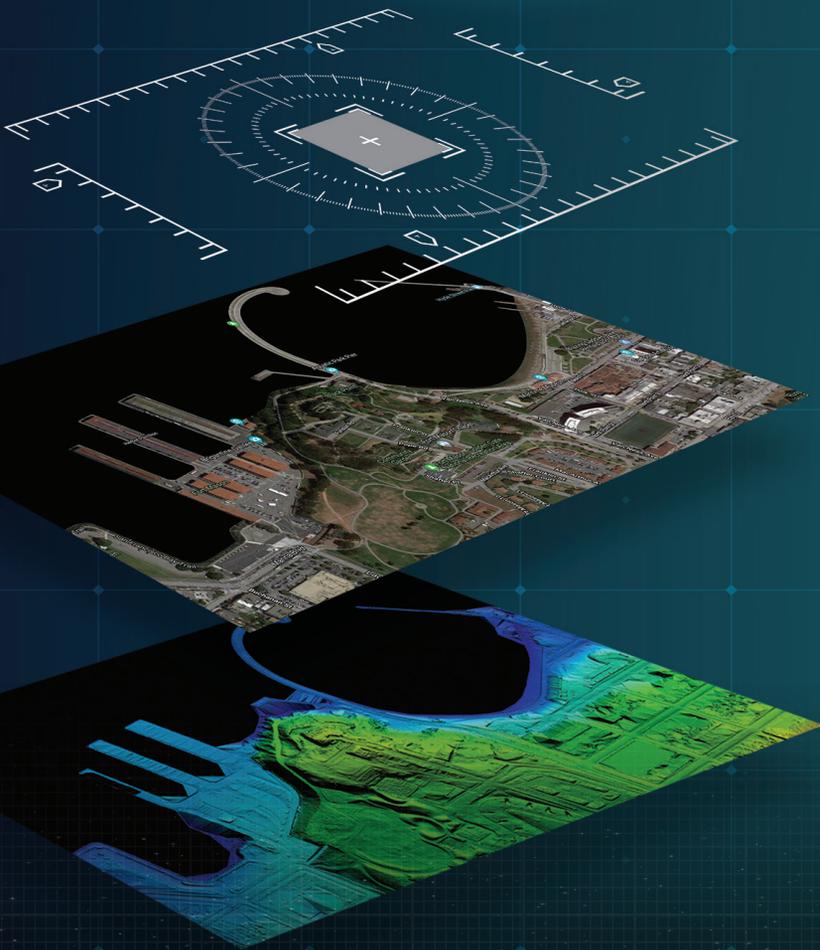
asprs.org

ASPRS

AT GEOMNI, WE MAP FOR SUCCESS



When you have Geomni collect imagery and data for your next project you will be harnessing the power of one of the largest Geo-spatial companies in the United States. With 14 hubs nationally, we are local to your next planned job site. Helping you map for success is our mission. We achieve that goal by providing access to our extensive data library, and offering custom imaging flights, while our fleet provides the regional knowledge, resolution, accuracy, and timeliness to deliver the critical information precision mappers need, at a price point fit for your project budgets.



HAWAII

NORCAL

SOCAL

OREGON

IDAHO

HEADQUARTERS

ARIZONA

NEW MEXICO

COLORADO

TEXAS

KANSAS

INDIANA

FLORIDA

SOUTH CAROLINA

MARYLAND

MAINE



WWW.GEOMNI.COM/PSM