

PE&RS

March 2019

Volume 85, Number 3

The official journal for imaging and geospatial information science and technology

PHOTOGRAMMETRIC ENGINEERING & REMOTE SENSING

Foxe Basin

Ice



PECORA 21 • ISRSE 38

Continuous Monitoring of Our Changing Planet: From Sensors to Decisions



Save the Date

October 6-11, 2019

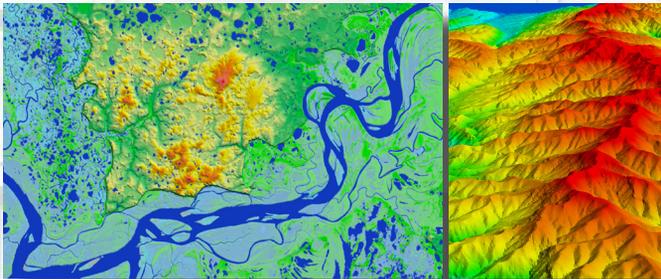
Baltimore, MD • Marriott Waterfront

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



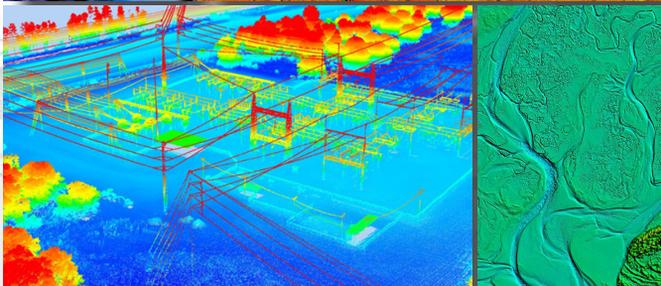
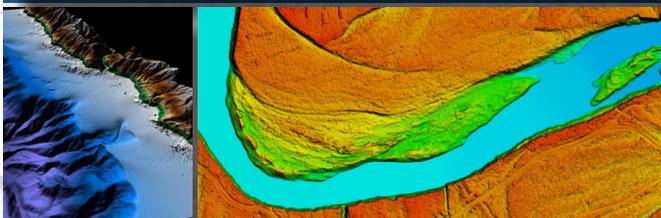
NEW FROM ASPRS AND DEWBERRY

DEM Users Manual, 3rd Edition



Digital Elevation Model Technologies and Applications: The DEM Users Manual, 3rd Edition

Edited by David F. Maune, PhD, CP
and Amar Nayegandhi, CP, CMS



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

Hardcopy		eBook ¹	
List Price	\$100	List Price	\$85
ASPRS Member	\$80	ASPRS Member	\$65
Student	\$50	Student	\$50

*Plus shipping/handling

¹Coming soon

The 3rd edition of the DEM Users Manual includes 15 chapters and three appendices. References in the eBook version are hyperlinked. Chapter and appendix titles include:

1. Introduction to DEMs
2. Vertical Datums
3. Standards, Guidelines & Specifications
4. The National Elevation Dataset (NED)
5. The 3D Elevation Program (3DEP)
6. Photogrammetry
7. IfSAR
8. Airborne Topographic Lidar
9. Lidar Data Processing
10. Airborne Lidar Bathymetry
11. Sonar
12. Enabling Technologies
13. DEM User Applications
14. DEM User Requirements & Benefits
15. Quality Assessment of Elevation Data
 - A. Acronyms
 - B. Definitions
 - C. Sample Datasets

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.

Chapters 1, 3, 5, 8, 9, 13, 14, and 15 are "must-read" chapters for users and providers of topographic lidar data. Chapter 8 addresses linear mode, single photon and Geiger mode lidar technologies, and Chapter 10 addresses the latest in topobathymetric lidar. The remaining chapters are either relevant to all DEM technologies or address alternative technologies including photogrammetry, IfSAR, and sonar.

As demonstrated by the figures selected for the front cover of this manual, readers will recognize the editors' vision for the future – a 3D Nation that seamlessly merges topographic and bathymetric data from the tops of the mountains, beneath rivers and lakes, to the depths of the sea.

Co-Editors
David F. Maune, PhD, CP and Amar Nayegandhi, CP, CMS

ORDER ONLINE
[HTTPS://WWW.ASPRS.ORG](https://www.asprs.org)



ANNOUNCEMENTS

G-Squared, LLC, a Fayetteville, Tennessee based mapping firm has completed 20 years of quality service to its clients. Principal and ASPRS Certified Photogrammetrist Lindsey Galyen, Jr. formed G-Squared in January 1999 with a small team of likeminded professionals for the sole purpose of producing highly accurate maps. The firm has remained dedicated to this task and continues to be an active member of the professional community.

G-Squared provides a wide range of mapping services including lidar acquisition and processing, digital aerial photography, airborne and ground GPS survey, digital orthophotography, engineering-grade topographic mapping, planimetric data updates, image processing and GIS for its clients. G-Squared performs engineering grade mapping and geodetic products that exceed industry standards.

G-Squared's client list includes engineering firms, counties, municipalities, airports, utilities, various departments of transportation, and federal agencies. They have extensive experience working with the State of Alabama and the Tennessee Department of Transportation.

Mr. Galyen attributes G-Squared's success to its reputation for quality products and client satisfaction. "We have built long-term, sustaining relationships with our clients. I attribute this to the time and effort our entire staff dedicates to fully understand our clients' requirements and ensure the project meets their needs."

Mr. Galyen said G-Squared performs all production at its 5,000 square foot facility at 20 Ardmore Highway, Fayetteville, Tennessee "It is important for our economy to keep high-tech jobs in our local communities," stated Galyen. "All production work is performed in Fayetteville, Tennessee by our technicians and professional staff who have extensive experience producing quality maps."

EVENTS

Commercial UAV Expo Europe takes place 8-10 April at the Amsterdam RAI, with pre-conference programming and workshops on Monday 8 April followed by conference programming and exhibits on Tuesday 9 April and Wednesday 10 April.

The event draws vertical industry end-users from across Europe, Asia, Australia the Middle East and around the world. Already, top professionals from a number of leading organizations have registered to attend.

In addition to the exhibits, Commercial UAV Expo Europe has extensive programming, offering both general sessions and breakout sessions that delve deeply into UAS for specific vertical industries. Vertical industry sessions focus on construction; energy & utilities; precision agriculture & forestry; public safety; surveying & mapping; transportation and infrastructure. End-user roundtables are being offered on four of these topics and are by invitation only. These will provide participants a deep dive into best practices, opportunities and threats when it comes to UAS in their respective sectors. "Not only is the content critical to these professionals, but the connections forged are priceless," said Ms. Murray. Additional features of the programming are vendor-delivered Product Previews, the 2019 Drone Hero Europe contest and Exhibit Hall Theater Programming.

Commercial UAV Expo Europe, presented by Commercial UAV News, is an international conference and trade fair exclusively focused on the commercial UAS market covering industries including Construction; Energy & Utilities; Forestry & Agriculture; Infrastructure & Transportation; Public Safety & Emergency Services; Security; and Surveying & Mapping. It takes place 8-10 April at the RAI Amsterdam. Its North American sister event takes place 28-30 October 2019 at Westgate in Las Vegas. For more information, visit www.expouav.com/europe and www.expouav.com.

CALENDAR

- 13 March, **Introductory Webinar—Using Earth Observations to Monitor Water Budgets for River Basin Management**. For more information, visit <https://arset.gsfc.nasa.gov/water/webinars/2019-river-basin>.
- 3-7 April, **AAG 2019 Annual Meeting**, Washington, DC. For more information, visit <http://annualmeeting.aag.org/>.
- 29-30 April, **Digital Image & Signal Processing**, Oxford, United Kingdom. For more information, visit www.disp-conference.org.
- 3-5 May, **5th International Conference on Geographical Information Systems Theory, Applications and Management**, Heraklion, Crete, Greece. For more information, visit <http://www.gistam.org>.

425 Barlow Place, Suite 210, Bethesda, MD 20814
301-493-0290, 225-408-4422 (fax)
www.asprs.org

Membership/PE&RS Subscription/Conferences

Priscilla Weeks — pweeks@asprs.org

Advertising/Exhibit Sales

Bill Spilman — bill@innovativemediasolutions.com

Peer-Review Article Submission

Alper Yilmaz — PERSEditor@asprs.org

Highlight Article Submission

Jie Shan — jshan@ecn.purdue.edu

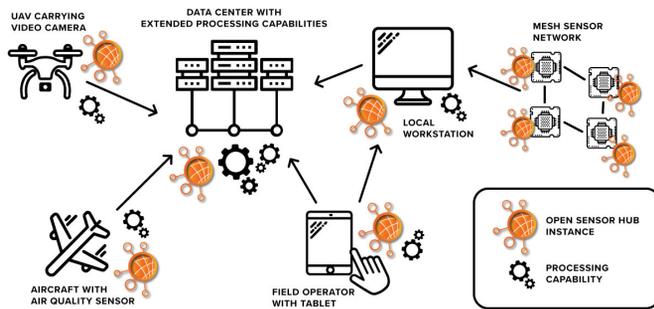
Industry News

Rae Kelley — rkelley@asprs.org

Calendar

calendar@asprs.org

FEATURES



157 SectorInsights.org— Remote Sensing, Drones, Unstuffed Systems (UxS), Sensors, and Interoperability at Open
 By Mark Reichardt, and Scott Simmons

COLUMNS

- 161** Book Review—*Imagery and GIS: Best Practices for Extracting Information from Imagery*
- 163** Grids and Datums
This month we look at the Republic of Angola.

ANNOUNCEMENTS

- 160** ASPRS Reviewers
- 162** ASPRS Rising Stars
- 167** Headquarters News
- 168** New ASPRS Members
Join us in welcoming our newest members to ASPRS.

DEPARTMENTS

- 154** Industry News
- 154** Calendar
- 159** Ad Index
- 196** ASPRS Sustaining Members

PEER-REVIEWED ARTICLES

169 An Image-Pyramid-Based Raster-to-Vector Conversion (IPBRTVC) Framework for Consecutive-Scale Cartography and Synchronized Generalization of Classic Objects

Chang Li, Xiaojuan Liu, and Wei Lu

Key problems in raster-to-vector conversion and cartographic generalization are addressed with an innovative image-pyramid-based raster-to-vector conversion (IPBRTVC) framework with quality control for consecutive-scale cartography and synchronized generalization.

179 Land Cover Classification in Combined Elevation and Optical Images Supported by OSM Data, Mixed-level Features, and Non-local Optimization Algorithms

Dimitri Bulatov, Gisela Häufel, Lukas Lucks, and Melanie Pohl

Shadows and strong intra-class variations of appearances make land cover classification from airborne data difficult in urban terrain. This paper proposes an algorithm using Open Street Map data, superpixel decomposition and a non-local optimization module operating on Markov Random Fields.

197 Improved Camera Distortion Correction and Depth Estimation for Lenslet Light Field Camera

Changkun Yang, Zhaoqin Liu, Kaichang Di, Yexin Wang, and Man Peng

Two improved methods of distortion model optimization and depth estimation are proposed for a light field camera. Experimental results show a 30% decrease in projection error and a 42% decrease in depth error compared with state of art algorithms.

209 Analysis and Correction of Digital Elevation Models for Plain Areas

Cristian Guevara Ochoa, Luis Vives, Erik Zimmermann, Ignacio Masson, Luisa Fajardo, and Carlos Scioli

DEMs created from ASTER, STRM, and ALOS PALSAR were used to model surface water movement in plain areas, with SRTM shown to produce the best results when IDW and ANUDEM interpolation methods are applied.

221 An Evaluation of Reflectance Calibration Methods for UAV Spectral Imagery

Jarrod Edwards, John Anderson, William Shuart, and Jason Woolard

This study demonstrates the need for standardized calibration procedures in the spectral correction of small-format remote sensor data by comparing AIRINOV and Empirical Line Calibration (ELC) method applied to Parrot Sequoia imagery.

231 Interaction Relationship Between Built-Up Land Expansion and Demographic-Social-Economic Urbanization in Shanghai-Hangzhou Bay Metropolitan Region of Eastern China

Rui Xiao, Xin Huang, Weixuan Yu, Meng Lin, and Zhonghao Zhang

Employing coupling coordination degree model (CCDM), this research attempts to reveal the interaction relationships between built-up land expansion (BLE) and demographic-social-economic (DSE) urbanization in Shanghai-Hang-zhou Bay (SHB) Metropolitan Region.

See the Cover Description on Page 156

COVER DESCRIPTION



Throughout most of the year, the waters of Foxe Basin are choked with sea ice. By the end of summer, however, open water typically dominates this part of the Canadian Arctic. That was the case when these images were acquired in September 2018, as small patches of ice lingered in the northern reaches of Hudson Bay around Prince Charles Island and Baffin Island.

The wide view was acquired by the Moderate Resolution Imaging Spectroradiometer (MODIS) on the Aqua satellite on 3 September 2018. The Operational Land Imager (OLI) on Landsat-8 acquired the detailed view (cover image) on 2 September. Notice in the wide view that the clouds appear whiter than the ice. The detailed image shows an even better view of this discoloration.

There are a number of reasons why ice can take on a brown tinge. Particles from natural and human sources—such as aerosols from industrial plants and ship emissions, or mineral dust from land—can blow in. Smoke particles from fires—such as those burning in Siberia in early July—also stream over the sea ice in the Arctic Ocean. If these particles settle onto the ice, they can darken the surface and increase melting.

Airborne sources, however, are probably not the reason for the brown ice in these images. The Foxe Basin is known for sea ice that gets stained brown by sediment from the surrounding land or from the shallow seafloor. Greg McCullough of the University of Manitoba points out that some of the color could also be caused by algae, which can grow under the ice and wash up onto the surface during a storm.

Tidal currents and winds can move the sea ice around and organize it into various patterns and tendrils. According to Jennifer Lukovich, also of the University of Manitoba, the sea ice in this image shows a signature of cyclonic sea ice circulation southwest of Prince Charles Island.

For more information, visit <https://landsat.visibleearth.nasa.gov/view.php?id=92764>.

NASA Earth Observatory images by Lauren Dauphin, using MODIS data from LANCE/EOSDIS Rapid Response and Landsat data from the U.S. Geological Survey. Story by Kathryn Hansen.

PHOTOGRAMMETRIC ENGINEERING & REMOTE SENSING



JOURNAL STAFF

Publisher ASPRS
Editor-In-Chief Alper Yilmaz
Assistant Editor Jie Shan
Managing Editor Melissa J. Porterfield
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.



By Mark Reichardt, and Scott Simmons

Remote Sensing, Drones, Unstaffed Systems (UxS), Sensors, and Interoperability at Open Geospatial Consortium (OGC)

Introduction

UxS, or unstaffed systems are fast becoming the norm for accomplishing time-critical observations, situational awareness, and material delivery. The myriad of airborne, ground-based, and water-borne UxS have a wide range of operational, data processing, and data management tasks tailored to support the end user – whether they are hobbyists or experts.

The ability to rapidly deploy and re-use key UxS workflows is a challenge, given the variety of platforms, observing sensors, and end-user requirements. To address this challenge, government and industry bodies are working to identify common interoperability arrangements that could streamline UxS deployment and maximize choice in the marketplace. This article discusses today's open standards environment as applicable to UxS interoperability. We will emphasize the context of location, which permeates all aspects of UxS operations: from mission planning, to capturing and processing of observations, to delivering a package to a doorstep.

Current Drone Architecture Environment

In the context of Unstaffed Aircraft Systems (UAS) or drones, many people refer to the current environment as the “Wild West.” Platform and sensor technologies are rapidly evolving, and new use-cases are being realized on a daily basis. The proliferation of relatively low-cost professional and consumer / hobbyist aircraft into the services marketplace has been particularly transformative.

The ‘consumerization’ of the technology means that UAS no longer require extensive training to deploy and operate, largely due to: (1) vast improvements in consumer-grade hardware miniaturization and (2) improved software to assist in-flight operation.

Opportunity exists now to align the UAS community on a common framework of open voluntary consensus standards to maintain a diverse marketplace, reduce vendor lock-in, and facilitate the rapid mobilization of UAS assets even as mission requirements and technologies change. The American National Standards Institute Unmanned Aircraft Systems Collaborative¹ with involvement of over 180 government,

industry organizations, associations and standards organizations, including OGC, is working to define an inventory of existing UAS voluntary consensus open standards, and to establish a roadmap of areas where standardization challenges still exist. This is an excellent start.

Of particular importance is the ability to use standards as part of architectures or best practices that can meet a set of common mission requirements. One important area involves using UAS-based remote sensing to meet a range of user needs, including urban planning, infrastructure inspection (i.e., pipeline, electric transmission, bridges, undersea cables), search and rescue, precision agriculture, and land administration, to name but a few.

Addressing Interoperability

A proven and widely implemented geospatial interoperability framework already exists to support general UAS operations (e.g., geofencing, flight planning) as well as imaging, point cloud, and other remote sensing observations from UAS' on-board sensors (Figure 1), and to process this information to support applications and decision making. The key is the ability of an open standards-based architecture to rapidly enable: the introduction of a range of observation platforms; the connection to, and integration of, multiple sensors; the management and processing of sensor observations; and, ultimately, the application of the sensor observations in a situational awareness or decision context.

Standards-based Architecture

The existing architecture for drone operations builds upon legacy concepts from general aviation and large UAS capabilities. Architectural elements can be divided into three general aspects of UAS use: (1) pre-flight mission planning and

Photogrammetric Engineering & Remote Sensing
Vol. 85, No. 3, March 2019, pp. 157–159.

0099-1112/18/157–159

© 2019 American Society for Photogrammetry
and Remote Sensing

doi: 10.14358/PERS.85.3.157

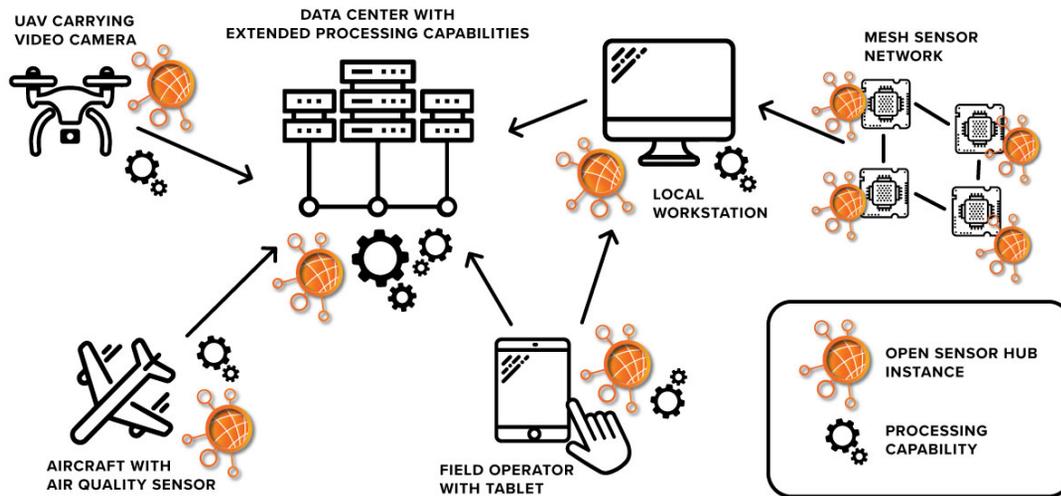
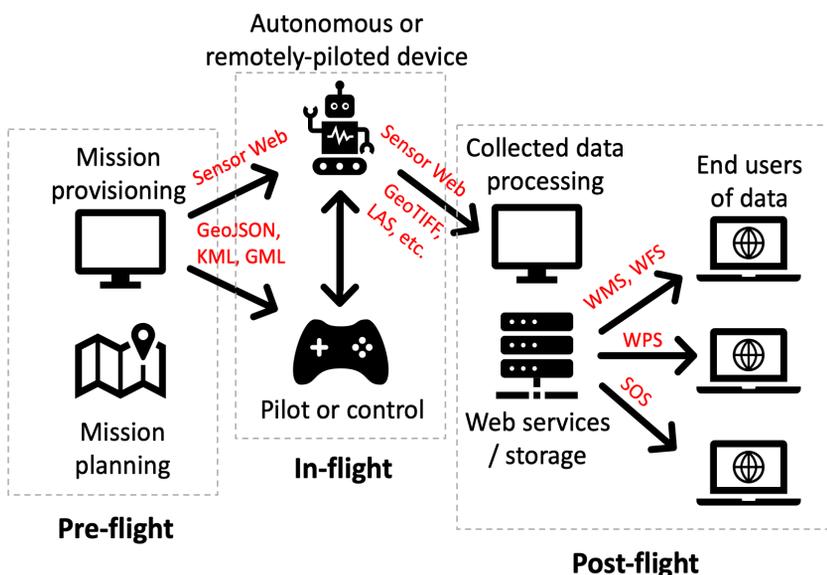


Figure 1. OpenSensorHub general architecture. Courtesy OpenSensorHub.

provisioning; (2) in-flight operation and command & control; and (3) post-flight data processing and dissemination. This architecture is best described in the context of an operational construct – from pre-flight to post-flight, as shown in Figure 2.

This architecture can also support Unmanned Traffic Management (UTM), facilitated in part by the Aeronautical Information Exchange Model (AIXM), which is based on OGC’s Geography Markup Language (GML). AIXM was developed by the US Federal Aviation Administration (FAA), the US National Geospatial Intelligence Agency (NGIA) and the European Organization for the Safety of Air Navigation (EUROCONTROL).

Pre-flight – mission planning and operational environment
 UAS operations occur over a specific geography, which the UAS needs to be able to navigate. Further, UAS missions are typically planned using other geographic information, such as boundaries, micro-climate, and/or other obstacles. It is critical, then, that all source information is correctly positioned with respect to the planned mission. There is a need to use standardized coordinate reference systems (CRSs) and data encodings that support those CRSs. Mission planning tools should support the integration of different data sources and provide flight parameters aligned with the GNSS (Global Navigation Satellite System) positioning of the UAS.



“OGC welcomes anyone, especially ASPRS members, interested in advancing solutions to these and other location-related issues...”

Figure 2. Standards-based interfaces between operational phases. Most of the standards listed in RED are from the OGC and more information on those standards are detailed in OGC Implementation Standards.² Geography Markup Language (GML), Web Map Service (WMS), Web Feature Service (WFS), Web Processing Service (WPS), Sensor Observation Service (SOS).

Once planned, mission details must be shared with the aircraft and/or operator. One benefit of using standardized encodings is the ability to share this content across multiple platforms. Given that many UAS operators use multiple aircraft from different manufacturers, the ability to use a single, standardized tool or workflow to mission-plan for an entire fleet of UAS is a massive boon. Such a tool is only possible when a common standards-based format, such as OGC GML or GeoJSON Geographic JavaScript Object Notation), is used for exchanging flight-plans.

Similarly, restricted airspaces must also be delivered as positionally-accurate geofences. Geofences can be defined in two or three dimensions and have temporal extents, either being valid for only certain times or moving with time. As with flight plans, geofences should be available in a shareable, standardized, and unequivocally-positioned format.

In-flight – flying, sensing, and communicating

UAS operations, whether by remote piloting or autonomous control, are managed in much the same architectural paradigm. Even with remotely-piloted aircraft, some adjustments to flight characteristics to compensate for local conditions may occur autonomously, and the implementation of these capabilities is best left to the hardware manufacturer. However, the actual issue of instructions, return of feedback, and adjustment of the operation is the essence of flight control.

UAS airframes and payloads include lightweight, low-power sensors that are best connected to the overarching command & control systems by IoT (Internet of Things) protocols. OGC's Sensor Web Enablement (SWE) standards, and the related SensorThings API standard³ facilitate development of a networked environment for the operation, data collection, data dissemination, processing, and application of data captured from UAS – or any form of UxS. Such an open standards-based environment has been demonstrated by numerous implementations, including the OpenSensorHub⁴ as shown in Figure 1.

Post-flight – logging activity and using the data

The processing and dissemination of data can begin in-flight or upon completion of the mission. Sensor payloads on the aircraft may collect imagery, lidar point clouds, or some other type of measurement. Ideally, the raw data includes some positioning information and is made available in formats such as GeoTIFF or LAS.⁵ The sensor data is combined with the aircraft flight logs and operational data, and manipulated and readied for dissemination. OGC web service standards, such as Web Map Service or Web Feature Service, provide the processed data and any derivative works to web-connected clients for further use.

What Next? Engaging the Community

While a strong set of useful standards are available to support UxS activities, there are some key areas that require additional focus by the technology, policy, and standards communities. Of particular interest to OGC are topics such as privacy, security, and unattended or autonomous operations. Another area of interest is weather. Today's weather services support civil and commercial airliner scale, but micro-climate scale weather information will be required to best fulfill UAS missions, specifications, and tolerances. Location accuracy is also necessary for collision avoidance and to deliver observations and material to the intended point of interest.

OGC welcomes anyone, especially ASPRS members, interested in advancing solutions to these and other location-related issues as we work with partners in the UxS community to advance best practices for the safe, efficient, and effective operation of unstaffed systems.

Authors

Mark Reichardt (mreichardt@myogc.org) is President and Chief Executive Officer of the Open Geospatial Consortium (OGC). Mr. Reichardt has responsibility for implementation of Consortium strategic goals and objectives, overseeing the development and promotion of OGC standards, and working to ensure that OGC programs foster member success in addressing global interoperability needs.

Scott Simmons (ssimmons@opengeospatial.org) is the Executive Director of the OGC Standards Program. In this role, he coordinates member-driven standards development activities and ensures that all OGC standards progress through the organization's consensus process to approval and publication.

1 ANSI: American National Standards Institute Unmanned Aircraft Systems Collaborative. www.ansi.org/standards_activities/standards_boards_panels/uassc/overview. Accessed 31 January 2019.

2 OGC Implementation Standard. 2019. <http://www.opengeospatial.org/docs/is>. Accessed 8 February 2019.

3 Liang S., C.-Y. Huang and T. Khalafbeige. 2016. OGC: OGC SensorThings API Part 1: Sensing. <http://docs.opengeospatial.org/is/15-078r6/15-078r6.html>. Accessed 31 January 2019.

4 OpenSensorHub. www.opensensorhub.org. Accessed 31 January 2019.

5 The ASPRS LAS Working Group, a part of the ASPRS Lidar Division, has released LAS 1.4 Revision R14 for Public Comment. [see <https://github.com/ASPRSorg/LAS/pull/76>. Accessed 8 February 2019].

Ad Index

Geomni, Inc.

Geomni.net/psm

Cover 4

ASPRS, the society, thanks all of the peer-reviewers, who make *PE&RS*, the journal, possible. Peer-Review is often the least rewarded and most important part of research publishing. Years of dedication to the exchange of ideas and the betterment of others continues to grow our robust research community. Their dedication and support has made *PE&RS* a flagship of the industry.

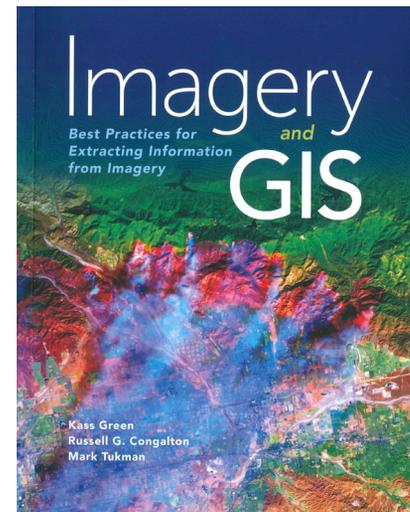
Congratulations, the 2018 *PE&RS* peer-reviewers:

Amr Abd-Elrahman	Ronan Fablet	Hao Li	Sarah Sutton
Abureli Aierken	Mst Ilme Faridatul	Huapeng Li	Yu Tao
Davood Akbarl	Ruyi Feng	Jiayi Li	Pardhasaradhi Teluguntla
Devrim Akca	Juan Flores	Mengmeng Li	Isabella Toschi
Selim Aksoy	Clive Fraser	Peijun Li	Charles Toth
Dimitrios D. Alexakis	Mauricio Galo	Peng Li	Damla Uca Avci
Ugur Alganci	Wenxia Gan	Qingyu Li	Bruno Vallet
Amin Alizadeh Naeini	Caroline M. Gevaert	John Liames	Yury Vizilter
Shrinidhi Ambinakudige	Suchi Gopal	Wenzhi Liao	Michele Volpi
Reda Amer	Anne Grote	Yu-Ching Lin	Chengyi Wang
Nan An	Cyrielle Guérin	Xiao Ling	Dongdong Wang
Ru An	Eberhard Gülch	Hui Liu	Hongshuo Wang
John Anderson	Murali Gumma	Shijie Liu	Mi Wang
Itiya Aneece	Qinghua Guo	Tao Liu	Xin Wang
Hasi Bagan	Xian Guo	Bing Lu	Timothy Warner
Nicolas Baghdadi	Xulin Guo	Qi Lv	Martin Weinmann
Gabor Barsai	Barry Haack	Ye Lyu	Jianguang Wen
William Baugh	Mostapha Harb	Lei Ma	Guisong Xia
Ilhami Bayramin	Preston Hartzell	Mathew Maimaitiyiming	Shaobo Xia
Paul Besl	Hadiseh Hasani	Gottfried Mandlbürger	Pengfeng Xiao
George Bevan	Wei He	Paul Manley	Linfu Xie
Francesca Bovolo	Yin He	J. McClinton	Lin Yan
Michaela Buenemann	Yuhong He	Jochen Meidow	Jie Yang
Dimitri Bulatov	Petra Helmholz	Adel Moussa	Xiguang Yang
Michael Campbell	Han Hu	Marco Neubert	Somayah Yavari
Nicolas Champion	Hai Huang	Ali Ok	Alper Yilmaz
Ting Chan	Xianfeng Huang	Sidike Paheding	Xiao Yong
Xiao Changlin	Xin Huang	Christopher Parrish	Qiangqiang Yuan
Bin Chen	Yao Huang	Norbert Pfeifer	Xiuxiao Yuan
Hujiaun Chen	Ming-Chih Hung	Rongjun Qin	Sebastian Zambanini
Liqiong Chen	Dorota Iwaszczuk	Mahesh Rao	Caiyun Zhang
Qi Chen	Salahuddin Jaber	Jiann-Yeou Rau	Qian Zhang
Lloyd Coulter	Karsten Jacobsen	Fabio Remondino	Su Zhang
Michael Cramer	Huiran Jin	Moritz Rexer	Tao Zhang
Nicolas David	M Kalamuthumari	Megan Ritelli	Xiaoyang Zhang
Philip Dennison	Van R. Kane	Petri Ronnholm	Xueliang Zhang
Sagar Deshpande	Joseph Knight	Umut Sefercik	Zhenchao Zhang
Kaichang Di	Mila Koeva	Di Shi	Jiaqi Zhao
Cenk Dönmez	Loïc Landrieu	Xiaopeng Song	Mei Zhou
Yongming Du	Chang Li	Rei Sonobe	Xiaolin Zhu
Almabrok Essa	Guoyuan Li	Douglas Stow	Zerun Zhu

Imagery and GIS: Best Practices for Extracting Information from Imagery is a new remote sensing textbook aimed at GIS users interested in remote sensing as a data source and rather than the technical details of creating remote sensing data. In the current state of remote sensing, where data are widely accessible, and geometric and radiometric pre-processing are often automated, this book is well suited to the majority of potential remote sensing users interested in the application of remote sensing data rather than the technical processes behind remote sensing. The book is organized into 14 chapters divided into 4 sections – “Discovering Imagery,” “Using Imagery,” “Extracting Information from Imagery,” and “Managing Imagery and GIS Data,” with acronyms, glossary, references, and index at the end of the book. Each chapter is organized with bold headings with an introduction that prepares the reader for why the information in this chapter is important and concludes with a summary that specifically considers “practical considerations” that GIS professionals using remote sensing will need to anticipate. Throughout the book, there are colorful and detailed tables, charts, and pictures that support the text, along with occasional references to ArcGIS software and products. The book also contains several real-world case studies that help connect concepts with practice.

Section 1 serves as an introduction to both the book, but also fundamental concepts such as types of GIS data structures, how imagery fits into GIS and the types of maps produced from imagery, the imagery workflow, and the fundamentals of remote sensing such as the electromagnetic spectrum, types of sensors, types of platforms, and characteristics of imagery in the spatial and spectral domains. This section of the book is detailed and up-to-date without being too technical or overwhelming to new students of remote sensing. I particularly like the inclusion of a dedicated section at the end of chapter 2 (Thinking About Imagery), that outlines image workflows in 4 steps helping the reader understand how the following sections support the ultimate goal of high-quality data from remote sensing. Section 1 concludes with an entire chapter dedicated to choosing the correct imagery with helpful figures and tables detailing current and future sensors and platforms. Although this list may become quickly dated, it serves as an excellent resource.

Section 2, “Using Imagery” is split into two chapters. “Working With Imagery” covers data formats, visual enhancements, filters, and mosaics, and how to access imagery in ArcMap or ArcGIS Online including incorporating remote sensing into story maps. “Imagery Processing: Controlling Unwanted Variation in the Imagery” directly refers back to the imagery workflow from section 1. This chapter does a particularly good job of describing the different factors that can affect radiometric and geometric characteristics of imagery without being too technical. Readers appreciate the sources and implications of radiometric and geometric corrections without getting into the



Imagery and GIS: Best Practices for Extracting Information from Imagery

Kass Green, Russell G. Congalton, Ph.D., and Mark Tukman

Esri Press. 2017. 436 pp, ISBN 9781589484542.

Reviewed by Benjamin W. Heumann, Ph.D., Assistant Professor, Department of Geography, Central Michigan University, Mt. Pleasant, Michigan.

trigonometry or physics of these corrections better suited for more advanced readers and textbooks.

Section 3, “Extracting Information,” is the largest section of the book covering 5 chapters and 158 pages. This section is both detailed and broad describing a wealth of techniques from conceptualizing classification schemes, digital elevation models, elements of image interpretation, data transformations, spectral pattern analysis, incorporation of ancillary geospatial data in data exploration, classifications (including object-based image analysis and machine learning) and concludes with image change analysis. While the previous sections were well structured and approachable, I found this sec-

Photogrammetric Engineering & Remote Sensing

Vol. 85, No. 3, March 2019, pp. 161–162.

0099-1112/18/161–162

© 2019 American Society for Photogrammetry and Remote Sensing

doi: 10.14358/PERS.85.3.162

tion of the textbook more difficult and challenging to use, as the chapter titles do not clearly indicate the topics covered in each chapter and several topics span multiple chapters. For example, the elements of image interpretation are part of chapter 9 “Data Exploration” which also includes topics such as Tasseled-Cap Transformation and DEM Hillshades, while the application of manual interpretation is part of chapter 10 “Image Classification.”

Section 4, “Managing Imagery and GIS Data” contain three chapters that address accuracy assessment and imagery data management including publishing/processing services, and a final concluding chapter. Of all the chapters, I found the “Summary – Practical Considerations” for accuracy assessment particularly well done encouraging the reader to think critically about both the implementation and interpretation of accuracy assessments. The chapter on “Managing and Serving Imagery” provides the strongest link between remote sensing and GIS and provides a useful technical guide on the various characteristics and tools that need to be setup and managed in order for remote sensing imagery and data to be fully incorporated into a Geographic Information System used by many users. For example, the authors detail mosaic datasets and how they can be used to deliver custom imagery to a user.

I appreciate the book’s approach of connecting remote sensing to GIS rather than as a stand-alone discipline. Given the growth of GIS over the last two decades and the increasing number of specialized GIS programs in colleges and universities, this book seems to fit in an important need of helping to train GIS professionals in remote sensing who may not become remote sensing specialists. The emphasis on best practices and practical considerations lends itself well to the GIS community who are focused on applications rather than development. The inclusion of limited references ArcGIS products appeals to GIS users, but the book makes an excellent textbook or reference no matter what software is being used.

However, there were a few missed opportunities that would have made this book truly excellent. First, the book focuses almost exclusively on the classification of vegetation, not only in the case studies, but also the methods presented. In the introduction to the textbook, the authors state the importance of remote sensing to a range of fields including precision agriculture, humanitarian aid, forestry, and mining. There is a missed opportunity to connect with the boarder GIS community for applications in these fields. Second, while the book is admirably up-to-date in regards to sensors, there is a notable lack of unmanned aerial system/ drone remote sensing. While drones are mentioned as a platform, the book is completely lacking in describing Structure from Motion and its products such as orthomosaics and digital surface models. Given the growth of drone remote sensing and its distinctly different characteristics and processing methods, this should be an essential part of any contemporary remote sensing textbook.

Imagery and GIS is a well-rounded and approachable introductory remote sensing textbook clearly written with GIS users and community in mind. The quality of the book both in content and printing (e.g., color figures and fonts) combined with an excellent glossary and index, all at an affordable price, make it a very appealing textbook or reference.



ASPRS
Rising Star
PROGRAM

Meet Our Rising Stars!

Chris Stayte
Sponsored by Woolpert, Inc.

Corey Ochsman
Sponsored by Woolpert, Inc.

Collin Hutcheson
Sponsored by Merrick & Company

Charles Krugger
Sponsored by Stahl Sheaffer Engineering, LLC

Kristine Taniguchi-Quan, Ph. D.
Sponsored by Pacific Southwest Region of ASPRS

The Rising Star program is focused on mentorship of tomorrow’s leaders within ASPRS and the geospatial and remote sensing community. Geospatial companies, agencies, and ASPRS regions can all benefit by participation in the program through the advancement and professional development of their employees or associates through the process of sponsorship.

For more information go to www.asprs.org/rising-star-program.



& GRIDS DATUMS

BY Clifford J. Mugnier, CP, CMS, FASPRS

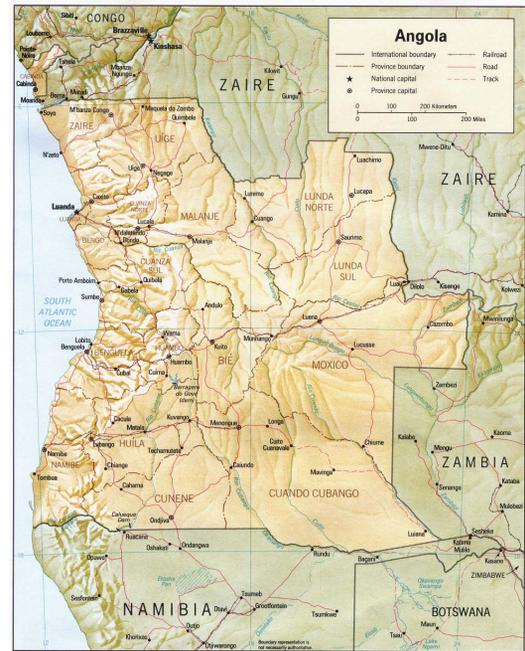
The Grids & Datums column has completed an exploration of every country on the Earth. For those who did not get to enjoy this world tour the first time, *PE&RS* is reprinting prior articles from the column. This month's article on the Republic of Angola was originally printed in 2001 but contains updates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

The original peoples of what is now Angola were probably Khoisan speaking hunters and gatherers (bushmen). During the first millennium AD, large-scale migrations of Bantu speaking people moved into the area and eventually became the dominant ethno-linguistic group of southern Africa. The most important Bantu kingdom in Angola was the Kongo, with its capital at Mbanza Kongo (called *São Salvador do Congo* by the Portuguese). South of the Kongo was the Ndongo kingdom of the Mbundu people. Angola got its name from the title for its king, the **ngola**. In 1483, Portuguese explorers reached Angola, Christianized the ruling family, and engaged in trade and missionary work. By the early 17th century, some 5,000+ slaves were being exported from Luanda annually. Angola received its independence from Portugal in 1975, but has been plagued by civil war and insurrections since independence. A familiar Bantu word in the U.S. is kwanza, which is Angola's unit of currency.

The interior forms part of the Central African Plateau, with elevations that range from 1,220 to 1,830 m (4,000 to 6,000 ft). The coastal plain is about 1,610 km long (1,000 miles) and varies in width from 48 to 160 km (30 to 100 miles). The highest point is Mt. Moco in the west at 2,559 m (8397 ft). The chief rivers include the Congo, Cuanza, and Cuene to the north, while south of the Lunda Divide some flow into the Zambezi River and others flow into the Okavango River.

Angola consists of two geographically separate expanses: Angola proper and Cabinda. Portuguese authority was not exercised continuously north of the Congo River in the present-day district of Cabinda until a relatively recent date. It was occupied by the Portuguese in 1783, but a French expedition forced them to evacuate the area 11 months later. Portugal laid definite claim to Cabinda in an additional convention to the Anglo-Portuguese treaty of January 22, 1815. Again, on February 26, 1884, an Anglo Portuguese treaty acknowledged claims by Portugal that included not only Cabinda and the Congo River inland to Nóqui but the whole Atlantic coast between 5° 12' and 8' South latitude. This produced a storm

THE REPUBLIC OF ANGOLA



of protests in Europe, and Portugal proposed a conference on the Congo that resulted in the Berlin Conference held between November 15, 1884 and February 26, 1885. Consider then, that the borders of Cabinda are in common with Congo (Brazzaville), which was part of the former Congo Français (French Congo), and is currently the Republic of the Congo. The southern border is in common with Congo (Kinshasa), once the Belgian Congo, later called Congo, then Zaire, and currently the Democratic Republic of Congo. The controlling classical datum for southern Africa and most surrounding countries of Angola proper is the Arc Datum of 1950 whose point of origin is station Buffelsfontein where $\Phi_0 = 33^\circ 59' 32.00''$ South, $\Lambda_0 = 25^\circ 30' 44.622''$ East of Greenwich, and the azimuth from south to station Zuurburg is $\alpha_0 = 183^\circ 58' 15''$. The reference ellipsoid for the Arc 50 Datum is the Clarke 1880 where the semi-major axis $a = 6,378,249.145$ m and the

Photogrammetric Engineering & Remote Sensing
Vol. 85, No. 3, March 2019, pp. 163–166.

0099-1112/18/163–166

© 2019 American Society for Photogrammetry
and Remote Sensing

doi: 10.14358/PERS.85.3.166

reciprocal of flattening $1/f = 293.465$. Angola's southern border is with that country once called German Southwest Africa, and Namibia is the only country in the African continent to utilize the Bessel 1841 ellipsoid for its Schwarzeck Datum where, for Namibia, the semi-major axis $a = 6,377,483.865$ International meters and the semi-minor axis $b = 6,356,165.383$ International meters. Values actually used in Namibia are $a = 6,377,483.865$ legal meters and $1/f = 299.1528128$. The origin point is Schwarzeck, near Gobabis, where $\Phi_0 = 22^\circ 45' 35.820''$ South, and $\Lambda_0 = 18^\circ 40' 34.549''$ East of Greenwich. (Our Paul M. Hebert School of Law here at Louisiana State University is helping build a Law Library at the University of Namibia).

The earliest geodetic observation in Angola is to the 1884 meridional distance from the mid-Atlantic island of St. Helena to Baie dos Elefantes (Elephant Bay) in Angola and thence north to São-Paulo de Loanda (Luanda). This revealed a probable error between $6'$ and $6\frac{1}{2}'$ in longitude in the existing Portuguese charts. This error was again reported in 1888 in a "Hydrographic Note" using meridional distances from Cape Town Observatory to the two ports mentioned above and to Moçâmedes. In 1891, *ANNALES HYDROGRAPHIQUES* of the French Navy published the telegraphic determination of longitudes for three sites in Angola as determined by Commander Pullen of the Royal English Navy. Those determinations were São-Paulo de Loanda – at the pavilion slab of the Fort of San Miguel – where $\Phi_0 = 08^\circ 48' 24''$ South, $\Lambda_0 = 10^\circ 53' 05''$ East of Paris, Benguela – at the Bureau Télégraphique (Telegraph Office) – where $\Phi_0 = 12^\circ 34' 43''$ South, $\Lambda_0 = 11^\circ 03' 40''$ East of Paris, and Moçâmedes – at the pavilion slab of Ponta da Noronha – where $\Phi_0 = 12^\circ 34' 43''$ South and $\Phi_0 = 11^\circ 03' 40''$ East of Paris.

In 1918, the Portuguese authorities established a new position for Luanda Observatory and supplied details of the triangulation of that area. "A Missão Geográfica de Angola criada na ...," "The Geographic Mission of Angola was created in the province in 1921 to establish a geodetic net in order to provide for a cartographic survey of the territory; after more than three campaigns, when it was dissolved, the Mission had surveyed two arcs of triangulation – one from Benguela to Huambo and the other along the" (*meridian*) " $14^\circ 00'$ East of Greenwich from adjacent figures to Lubango. Subsequently, for geodetic operations, a hydrographic mission, which had campaigns from 1930 to 1933, established geodetic figures along the Zaire (Congo) River with common vertices with the Congo (Kinshasa) Triangulation, and executed a geodetic base with astronomical observations for latitude in eight stations, for longitude in five stations, and two for azimuth. Furthermore, the systematic and regular occupation of geodetic points in the province was initiated in the beginning of 1941 as an actual activity of the Geographic Mission of Angola.

The instructions were formulated by the "Junta de Investigações do Ultramar" (Overseas Research Commission) of which the following points, with respect to the geodetic activities, were

1. complete the geodetic coverage of the province, including the principal triangulation figures and the secondary triangulation, first to involve two arcs of the meridian and the parallel at a 2° interval, and second, to densify adequately for cartographic operations at comprehensive spaces in the principal triangulation figures;
2. establish geodetic bases at distances of about 400 km along principal figures, and provide for the distribution that forms a precision of European second-order triangulation – which is a precision less than attributed to the principal triangulation;
3. observe Laplace stations at 200-km by 200-km spacings;
4. obtain the following precision for the geodetic and astronomical observations as principal triangles $-6''$, secondary triangles $-10''$, latitude $-0.2''$, longitude $-0.5''$, and azimuth $-0.5''$." To paraphrase, the Chief of the Geographic Mission later sent instructions in 1946 to observe arcs of the meridian and the parallel at a multiple interval of 3° . In 1953, the first order triangulation was adjusted using tables from the U.S. Coast and Geodetic Survey (*based on the Clarke 1866 ellipsoid*). "This adjustment was based on the condition equations formed by 269 triangles."

The triangulation eventually comprised 17 bases at Qui-mongo, Congo Yala, Quitinda, Luanda, Lucala, Camacupa, Catumbela, Alto Catumbela, Cimo, Serpa Pinto, Quipungo, Moçâmedes, Humbe, Baia dos Tigres, Sare, Namatuco, and Luiana – determined 1722 points based on observations of 3809 triangles of which 856 points were not occupied – made astronomical observations of latitude, longitude, and azimuth at 15 station – and performed the leveling of 943 km.

Immediately prior to WWII, the Belgians published booklets in French and Flemish, listing the coordinates and geodetic positions for Bas-Congo that extended from the Atlantic Ocean to Leopoldville. In those booklets are some 38 points that are referred to the Congo-Yella Datum. Some of those points are Portuguese control points common with Angola that shares the river. The fundamental point for the Congo-Yella Datum is the east base terminal of Congo-Yella where $\Phi_0 = 06^\circ 00' 53.139''$ South, $\Lambda_0 = 12^\circ 58' 29.287''$ East of Greenwich, and the ellipsoid is Clarke 1880. Transformation parameters from the Congo-Yella Datum to the Angola Datum (also Clarke 1880) are $\Delta X = -35.08$ m, $\Delta Y = +184.83$ m, and $\Delta Z = +63.02$ m (± 3.7 m); and from Congo-Yella Datum to the Camacupa Datum of 1948 are $\Delta X = -44.47$ m, $\Delta Y = +179.47$ m, and $\Delta Z = +59.30$ m (± 2.0 m). The similarity of values gives confirmation that, in this region, the Angola Datum is equivalent to the Camacupa Datum of 1948. From the Congo-Yella Datum to the WGS 84 Datum, $\Delta X = -93.28$ m, $\Delta Y = -164.11$ m, and $\Delta Z = -169.02$ m (± 5 m). Grids based on the Congo-Yella Datum are the Belgian Congo Gauss-Krüger Transverse Mercator Fuseau 13 (Zone 13) where latitude of origin (ϕ_0) = $2^\circ 30'$ South, Central Meridian (λ_0) = 13° E, FN = 700 km, FE = 220 km, and the scale factor. (m_0) = 1.0; Fuseau 14 where latitude of origin (ϕ_0) = Equator, Central Meridian

(λ_0) = 14° E, FN = 10,000 km, FE = 500 km, and the scale factor (m_0) = 0.9999; and Fuseau 16 where latitude of origin (ϕ_0) = Equator, Central Meridian (λ_0) = 16° E, FN = 10,000 km, FE = 500 km, and the scale factor (m_0) = 0.9999. The Gabon Belt (Fuseau Gabon) is found in use along the northern coast (PE&RS, September 1998). Of course, the UTM grid is common in Angola.

The Camacupa Datum of 1948 is based on the origin at Campo de Aviação where $\Phi_0 = 12^\circ 01' 19.070''$ South, $\Lambda_0 = 17^\circ 27' 19.800''$ East of Greenwich, and $h_0 = 1508.3$ m. Thanks to John W. Hager, "This is the principal vertex marked by a concrete monument, constructed on a high part of the Camacupa Air Field, immediately to the north of the runways." (*This is*) "defined as the "Datum Point" of the main triangulation network of Angola. A concrete monument with the dimensions 70 × 60 × 100 cm (length E-W × width N-S × height), topped by a white marble slab on which is cut in black: M.G.A. -P.F.- 1948; in the center of which is placed the top mark of the base, which is defined as the extreme West of the Geodetic Base of Camacupa. The mark found here is protected by a masonry casing with a metallic cover, easily removed to permit observations over the base when necessary." Remarkably, some Datums established by the Portuguese in Angola (and Moçambique) were referenced to the Clarke 1866 ellipsoid (the same as used in the U.S. for the North American Datum of 1927) where $a = 6,378,206.4$ m and $b = 6,356,583.8$ m. The only transformation parameters I have ever been able to scrounge from this Datum to WGS84 were obtained from Prof. Charles L. Merry at the University of Cape Town where $\Delta X = -49$ m, $\Delta Y = -301$ m, and $\Delta Z = -181$ m; and Prof. Merry estimates the accuracy at ±60 meters. According to Hager, "sometime in the 1960s or 1970s, DMA was asked to put Angola, then on the Camacupa Clarke 1866 Datum, on the Camacupa Clarke 1880 and the Arc 50 Datums. The Portuguese provided all the coordinates based on the Clarke 1866 Datum. They also provided tables to convert from the Clarke 1866 Datum to the Clarke 1880 Datum assuming that the tangent point of the two ellipsoids was at Camacupa. I think that the tie was on the 12th Parallel South to the Zambian Triangulation. (*It was*). The 6th Parallel south and Bas Congo surveys of Congo occupy common points with the Angola surveys and were adjusted to the Arc Datum of 1950. A comparison of the Angola values showed that Arc 50 Datum in Angola was adequate for mapping purposes. Angola is on the UTM Grid. I did find a local grid for Luanda and would expect other similar ones. For Luanda, the 1:2,000-scale city map plots directly on top of the UTM Grid of the 1:100,000-scale map. The 50,000 50,000 intersection is, in UTM coordinates, N = 9,024,000 and E = 306,000. This then results in a local Grid, Transverse Mercator projection, Clarke 1880 ellipsoid, $\phi_0 = 0^\circ$, $\lambda_0 = 15^\circ$ E, FN = 1,026 km, FE = 244 km, and (m_0) = 0.9996. The UTM scale factor at local 50,000 50,000 is 1.00006581. A unity scale factor would be expected for a City Grid, and this is pretty close to unity. The math for the false coordinates is FN = 10,000,000 –

9,012,000 + 50,000 = 1,026,000 and FE = 500,000 – 306,000 + 50,000 = 244,000." The Camacupa Clarke 1880 Datum is oftentimes referred to by the hydrographic community as the Camacupa-Vumbatumba Datum of 1950 based on the origin surveyed by MHCA in 1950 as $\Phi_0 = 06^\circ 26' 17.111''$ South and $\Lambda_0 = 12^\circ 27' 22.978''$ East of Greenwich. Transformation parameters used by Western Geophysical from the Camacupa-Vumbatumba Datum of 1950 to the WGS 84 Datum are $\Delta X = -39.44$ m, $\Delta Y = -353.66$ m, and $\Delta Z = -224.16$ m, and the transformation parameters used by the British Navy are $\Delta X = -48.81$ m, $\Delta Y = -343.58$ m, $\Delta Z = -228.32$ m, ±10 meters for the northern part of the country.

Thanks to parameters published into the public domain by the European Petroleum Studies Group (EPSG) headed up by Mr. Roger Lott of British Petroleum, there are a number of transformations from the Clarke 1880 version of the Camacupa Datum of 1948. For instance, Camacupa 1948 to WGS 72BE: $\Delta X = -37.2$ m, $\Delta Y = -370.6$ m, and $\Delta Z = -228.5$ m; this was derived by Geophysical Services, Inc. in 1979. Camacupa 1948 to WGS84, used by Conoco for Offshore Block 5: $\Delta X = -42.01$ m, $\Delta Y = -332.21$ m, and $\Delta Z = -229.75$ m. Camacupa 1948 to WGS84 and used by Topnav at PAL F2, by Elf in blocks 3 and 17 since 1994, and by Total in block 2 since 1994: $\Delta X = -50.9$ m, $\Delta Y = -347.6$ m, and $\Delta Z = -231$ m. An additional eight versions of parameters are used for the "same" transformation in offshore areas spanning the entire coast of Angola.

The MHASt Datum of 1951 (Missão Hidrográfica de Angola e São Tomé) fundamental point is a concrete block, point Y, at Malongo lighthouse that is at $\Phi_0 = 05^\circ 23' 30.81''$ South, $\Lambda_0 = 12^\circ 12' 01.59''$ East of Greenwich, and is referenced to the International ellipsoid of 1924 where $a = 6,378,388$ m and $1/f = 297$. From MHASt to WGS84: $\Delta X = -252.95$ m, $\Delta Y = -4.11$ m, and $\Delta Z = -96.38$ m. The Malongo Datum of 1987 replaced the MHASt Datum of 1951, and is also referenced to the same fundamental point (new coordinates unknown). The same ellipsoid is used; however, the transformation parameters have changed to become Malongo 1987 Datum to WGS 84: $\Delta X = -254.10$ m, $\Delta Y = -5.36$ m, and $\Delta Z = -100.29$ m, thanks to Mal Jones of Perth, Australia.

Hager went on to say; "A survey was done across Congo (Kinshasa) connecting Angola proper to Cabinda but the data were destroyed by a fire in Lisbon so Cabinda is on a local datum. About all the booklet for Cabinda will say is that it is not on Camacupa 1948 Datum. The values of the boundary marks in the northwest of Cabinda are in agreement with those published by (the French) IGN and used by Congo (Brassaville)."

Other datums existing in Angola include the Lobito Datum of 1937 based on the origin point at the astronomical pillar Restinga do Lobito, Extremo NE da Base do Lobito, where $\Phi_0 = 12^\circ 19' 00.86''$ South, $\Lambda_0 = 13^\circ 34' 45.67''$ East of Greenwich, Clarke 1866 ellipsoid. Dr. José Carvalho of Maputo, Moçambique states that the Camacupa Datum of 1948 coordinates of the same point are $\Phi_0 = 12^\circ 19' 01.357''$ South, $\Lambda_0 = 13^\circ 34' 58.375''$ East of Greenwich. The transformation from the

Lobito 1937 Datum to the WGS 84 Datum is $\Delta X = -256.73$ m, $\Delta Y = 0.00$ m, and $\Delta Z = -103.67$ m (± 10 m).

The Luanda Datum is based on the origin point at Luanda Observatory where $\Phi_0 = 08^\circ 48' 46.8''$ South, $\Lambda_0 = 13^\circ 13' 21.8''$ East of Greenwich, Clarke 1866 ellipsoid. The Moçamedes Datum of 1956 origin point is at the Moçamedes Meteorological Station where $\Phi_0 = 15^\circ 11' 16.34''$ South, $\Lambda_0 = 12^\circ 07' 34.53''$ East of Greenwich, Clarke 1866 ellipsoid.

Many thanks to Sequoia Read of the Defence Geographic Centre and the Geodesy Section of the United Kingdom Hydrographic Office.

UPDATE

One recent article¹ found is by Sebastian J.F.², Kutushev S.B.³, in 2015, called Modernization of “Geodetic Net” of the Republic of Angola. The “Base Tasks.”⁴ Printed in the Russian journal *Geodesy and Cartography* [Geodezija i kartografija / Геодезия и Картография], 2, pp. 19-24. However, the article itself was not available (21JAN2019).

The article’s abstract reads:

The National Surveying and Mapping Organization needs to be provided with legal support and technical documentation. In general, modernization should involve three types of reference nets – Geodetic, Leveling and Gravimetric.

It seems there have been many calls for the modernization of Angola’s geodetic network as far a field as the United Nations Framework Convention on Climate Change⁵ to a variety of dissertations^{6,7} from Moscow State University.

The 2012 dissertation, outlines the main tasks of an organizational and technical nature, which, according to the authors, is necessary to solve when upgrading the cartographic-geodetic of the Republic of Angola, State Geodetic Survey

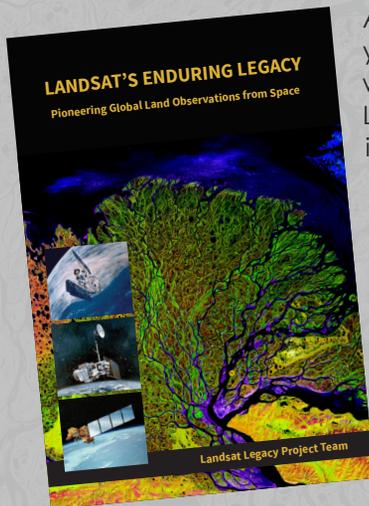
In a recent email note from Dr. Rui Fernandes,⁸ he mentioned that he inquired as to what the Angolan government had accomplished in terms of a new datum and coordinate system. Their reply was “We used your past project,” and they offered no further details (personal communication, January 2019).

- 1 *Geodesy and Cartography*. http://geocartography.ru/en/scientific_article/2015_2_19-24, Accessed 4 February 2019.
- 2 Department of Geography and Military Cartography of the General Staff Angola Armed Forces, Luanda, Angola.
- 3 Bashkir State University, Ufa, Russia.
- 4 DOI: 10.22389/0016-7126-2015-896-2-19-24.
- 5 Angola Initial National Communication <https://unfccc.int/resource/docs/natc/agonc1.pdf>, Accessed 4 February 2019.
- 6 Alves, Antonio. 2012. “Development of methods for upgrading the planned geodetic network of the city using modern satellite technologies”. <http://www.dissercat.com/content/razrabotka-metodiki-modernizatsii-planovoi-geodezicheskoi-seti-goroda-s-ispolzovaniem-sovrem>, Accessed 4 February 2019.
- 7 Lucian, Damiao. 1998. “Development of a project for the creation of the State Geodesic Network of Angola using satellite technologies.” <http://www.dissercat.com/content/razrabotka-proekta-sozdaniya-gosudarstvennoi-geodezicheskoi-seti-angoly-s-primeneniem-sputni#ixzz5eZrvGUUr>, Accessed 4 February 2019.
- 8 Universidade da Beira Interior Department of Computer Science SEGAL, Covilhã, Portugal.

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*.

Order online at amazon.com



After more than 15 years of research and writing, the Landsat Legacy Project Team is about to publish, in collaboration with the American Society for Photogrammetry and Remote Sensing (ASPRS), a seminal work on the nearly half-century of monitoring the Earth’s lands with Landsat.

Landsat Legacy Project Team

Samuel N. Goward, Darrel L. Williams, Terry Arvidson, Laura E. P. Rocchio, James R. Irons, Carol A. Russell, and Shaida S. Johnston



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 Helmholz, Ph.D., Petra.Helmholz@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_whu@163.com

Prasad Thenkabail, Ph.D., pthenkabail@usgs.gov

Ruisheng Wang, Ph.D., ruiswang@ucalgary.ca

Desheng Liu, Ph.D., liu.738@osu.edu

Valérie Gouet-Brunet, Ph.D., valerie.gouet@ign.fr

Yury Vizilter, Ph.D., viz@gosniias.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

Assistant Editor

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

Managing Editor

Melissa J. Porterfield, Ph.D., managingeditor@asprs.org

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

Amr Abd-Elrahman, Ph.D., sectorinsights@asprs.org

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

William C. Wright, Ph.D., William.Wright@USMA.edu

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

BE AN ASPRS MEMBER CHAMPION!

ASPRS is recruiting new members and **YOU** benefit from each new member **YOU** champion.

Not only can you contribute to the growth of ASPRS, but you can support the mission of ASPRS and spread the Member Value Proposition to your colleagues and students.

- ✓ For every new member you recruit, you will earn \$5 in ASPRS BUCK\$
- ✓ Recruit 5 new members and earn \$25 in ASPRS BUCK\$
- ✓ Recruit 10 or more new members and you will be honored with the Roger Hoffer Membership Award at the ASPRS Annual Conference, and one year of complimentary membership

ASPRS BUCK\$ VOUCHERS are worth \$5 each toward the purchase of publications or merchandise.

REMEMBER!

To receive credit for a new member, the CHAMPION'S name and ASPRS membership number must be included on the new member's application.

Contact Information

For Membership materials, contact us at:

301-493-0290, ext. 109 or 110

members@asprs.org

Those eligible to be invited to join ASPRS under the Member Champion Program are:

- Students and/or professionals
- Former ASPRS members are eligible for reinstatement as new members in this campaign

HELP ASPRS GROW BY PARTICIPATING IN THE MEMBER CHAMPION CAMPAIGN!

ASPRS MEMBERSHIP

ASPRS would like to welcome the following new members!

At Large

Chung Ryeol Hahn
Reza Karamooz
David J. McLelland
Chad Stewart Riggle

Central New York

Kris Hartley

Columbia River

Spencer Keller
Katharine Hannah Nicolato

Eastern Great Lakes

John Patrick May

Florida

Vanessa Alborno
Venous Babolhavaeji
Ryan W. Case
Youssef O. Kaddoura
Somayeh Mafi
Zachary Poole
Adam Rose
Oscar Rothrock

Heartland

Adam Clark
Jared Langley
James M. Powers

Mid-South

Joseph Fehrenbach
Erik S. Garcia
Benjamin H. Greenway
Stephen MacDonald
Ronald Okwemba
Tanner Polen
Anthony Bryan Ray

New England

Parinaz Rahimzadeh

Pacific Southwest

Kendell Hillis
Joseph R. Spann

Potomac

Shawn Michael Asher
Meher Bhagia
Michael Caldwell
Matt De La Rosa
Nia Meadows
Julie K. Nacos
Michael Tutwiler

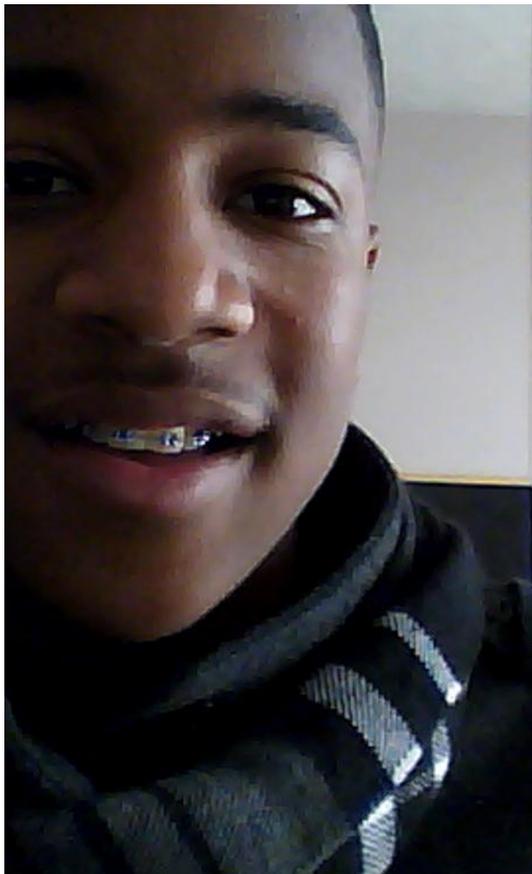
Rocky Mountain

William Gallery
Mary Garner
Aron Langley
Anna B. Pavlenko, Ph.D.
Jon Rohrer

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



Too young to drive the car? Perhaps!
But not too young to be curious about geospatial sciences.

The ASPRS Foundation was established to advance the understanding and use of spatial data for the betterment of humankind. The Foundation provides grants, scholarships, loans and other forms of aid to individuals or organizations pursuing knowledge of imaging and geospatial information science and technology, and their applications across the scientific, governmental, and commercial sectors.

Support the Foundation, because when he is ready so will we.

asprsfoundation.org/donate



An Image-Pyramid-Based Raster-to-Vector Conversion (IPBRTVC) Framework for Consecutive-Scale Cartography and Synchronized Generalization of Classic Objects

Chang Li, Xiaojuan Liu, and Wei Lu

Abstract

There are some key problems in raster-to-vector conversion and cartographic generalization, which include (1) deficient automation and low accuracy in the traditional raster-to-vector conversion processing; (2) data-source inconsistency in cartographic generation, i.e., different raster data sources converted to vector; and (3) how to acquire arbitrary-scale vector data. To solve these problems, we initially propose an innovative image-pyramid-based raster-to-vector conversion (IPBRTVC) framework with quality control for consecutive-scale cartography and synchronized generalization, of which details can be modified accordingly under the IPBRTVC framework. Landsat-8 imagery and Defense Meteorological Satellite Program (DMSP)/Operational Linescan System (OLS) night-time light imagery are used as a test dataset to extract classic objects in the geometry level. Experimental results show that the IPBRTVC framework not only solves the aforementioned problems well but also (1) improves efficiency of data processing by avoiding problems of corresponding features matching and topology errors, (2) contributes to develop relevant parallel computing system, and (3) helps to integrate the raster-to-vector conversion and consecutive-scale cartography.

Introduction

The scale issue is known as one of the fundamental scientific problems in remote sensing and cartography. The scale may have variability, that is, targets at different scales will show different characteristics (Wu and Li 2009). The scale effect issue has an important impact on the applicability of geographical laws and algorithms deduced and induced by different scales.

The scale issue of remote sensing data lies in the diversity of images with various spatial, spectral and temporal resolutions. For example, the quality of spatial resolution reflects the spatial details and the ability to extract the information of the scene itself (Woodcock and Strahler 1987). Zhao and Bo (2013) have enhanced the resolution of 500 m Moderate Resolution Imaging Spectroradiometer (MODIS) reflectance images and increased the spatial details of original MODIS images by using super-resolution via sparse representation, which helps

to solve the poor prediction for areas with a large number of mixed pixels. Visualization presentation for multi-scale remote sensing images plays a vital role in comprehensively understanding real land features by mapping the 3D world from a 2D image via remote sensing techniques. Using geographic object-based image analysis (GEOBIA) with very high spatial resolution (VHR) aerial imagery (0.3 m spatial resolution) to classify vegetation, channel and bare mud classes in a salt marsh, Kim *et al.* (2011) found that the multi-scale GEOBIA produced the highest classification accuracy, while the single-scale approach produced an only moderately accurate classification for all marsh classes. The image-pyramid-based method (Adelson *et al.* 1984), which is used to acquire multi-scale images with different resolutions by resampling and filter decomposition, is the major method to realize the visualization presentation for remote sensing images. Combining the pyramid transform and difference gradient detection algorithms is one method to generate multi-scale linear decomposition images (Simoncelli and Freeman 1995). It is practicable to improve the pyramid feature detection accuracy by means of bringing in steering coefficients to control the orientations of pyramid transform according to the study (Greenspan *et al.* 2002).

In cartography, the scale issue is seen as a main limiting factor (Weibel 1997). Graphic elements in different spatial scales are integrated with different mergeability and abstractability for properties. Li and Choi (2002) analyzed the association of the change in the number of symbols or the percentage of open space on similar scales with thematic attributes. They studied road features on topographic maps of Hong Kong from 1:1000 to 1:200 000 considering six types of thematic attributes, including “type”, “length”, “width”, “number of lanes”, “number of traffic ways” and “connectivity”. More progress has been made in the study of multi-scale representation in thematic maps, such as drainage maps. The U.S. Geological Survey (Brewer, Buttenfield, and Usery 2009), for instance, has achieved multi-scale (from 1:200 000 to 1:100 000 000) representation for national hydrological maps based on prior knowledge, such as terrain, and climate. Four of these experiments on basins show that corresponding prior knowledge has instructive influence on cartographic generalization. In terms of the scale transform, Buttenfield *et al.* (2010) has achieved multi-scale (from 1:500 000 to 1:200 000) hydrographical maps via cartographic generalization on the basis of the scale of 1:240 000, which succeeds in the transformation from

Chang Li and Wei Lu are with the Key Laboratory for Geographical Process Analysis & Simulation, Hubei Province, and College of Urban and Environmental Science, Central China Normal University, Wuhan, China, No. 152 Luoyu Road, Wuhan, Hubei, 430079, P.R. China (lcshaka@126.com (C.L.); liuxj58@mails.sysu.edu.cn (X. L.); and 13477463404@163.com (W.L.)).

Xiaojuan Liu is with School of Geography and Planning, Sun Yat-sen University, Guangzhou, China, No. 135 Xingang West Road, Guangzhou, Guangdong, 510275, P.R. China

Photogrammetric Engineering & Remote Sensing
Vol. 85, No. 3, March 2019, pp. 169–178.
0099-1112/18/169–178

© 2019 American Society for Photogrammetry
and Remote Sensing
doi: 10.14358/PERS.85.3.169

large-scale hydrographic maps to small-scale hydrographic maps. However, multi-scale representation that can only be achieved through a series of maps at fixed scales may result in a discontinuity, i.e., with jumps in the transformation between scales and a mismatch between the available scales and users' desired scales. Therefore, Li and Zhou (2012) proposed an integrated approach to build a hierarchical structure of road networks for continuous multi-scale representation purposes, with which the efficient and continuous multi-scale selective omission of road networks becomes feasible.

It is of great theoretical significance and practical value for organically integrating remote sensing data processing with cartographic generalization data processing and improving geographic data processing to apply relevant cartography methods and principles to raster-to-vector conversion. Both the thinning algorithm (Rosenfeld and Davis 1976) and centerline extraction method (Jimenez and Navalon 1982) are traditional in raster-to-vector conversion. The thinning algorithm easily generates noise at inflection points, intersections and branches, and the centerline extraction method easily ignores objects edges, so that they are both low-accuracy. Aiming at these problems, Hori *et al.* (1993) proposed a vectorization method by line fitting based on contours and skeletons. This method improved the accuracy of vectorization under a certain time complexity. K3M, a type of thinning method proposed by Saeed *et al.* (2010), improves and generalizes KMM (Saeed *et al.* 2001) and proves the universality—not only does it work for printed words, but it has also shown a remarkable success rate in thinning handwritten scripts, signatures, numbers, characters of non-Latin-based alphabets (Arabic, Japanese), and even biometric images such as faces or fingerprints, and attached great significance to the development of the thinning algorithm. In the area of cartographic generalization, studies on road centerline extraction or road vectorization are popular and significant. Ameri (2016) proposed a novel road vectorization methodology based on the image space clustering technique and weighted graph theory. Described in the paper (Yang Wei 2016), the method of road centerline extraction by a large volume of taxi global positioning system (GPS) trajectory data based on the Delaunay triangulation model, can improve the accuracy of road geometry and topology above 10% compared with conventional methods.

In conclusion, we need to focus on (1) the manual raster-to-vector conversion processing that mounts a work effort and easily makes deviations because of its dependence on the operator's skills and patience, which is why this traditional method has high labor and time costs but low accuracy; (2) the different datasource-based multi-scale presentation for geographical spatial data that increases the information processing time complexity and leads to spatial inconsistency problems in the application of raster and vector data; and (3) how to acquire arbitrary-scale vector data. Therefore, we aim for the raster-to-vector automatic conversion and consecutive-scale cartography based on the same remote sensing image data source via pyramid transformation.

The raster-to-vector conversion in this paper is in the geometry level without the property level, i.e. feature extraction and vectorization of classic object. Actually, raster-to-vector conversion refers to both semantic extraction and feature extraction. We take Environmental Systems Research Institute (ESRI), the most authoritative company in geographic information system (GIS), for example. (1) ArcGIS Help for the toolbox in ArcGIS has introduced how to work with the Raster-to-Polygon tool.¹ From the ArcGIS Help, we can find that raster field is optional, and the same is true for the raster-to-point and raster-to-polyline

tools, which means that raster-to-vector conversion not only includes high-level features (i.e. semantic extraction) but also low-level features (i.e. feature extraction). (2) Greenberg *et al.* (2005) performed a raster-to-vector conversion using ENVI 4.1 (ESRI), that has the same operating principle with the tool in ArcGIS, on the shadowed vegetation mask to generate a polygon coverage of contiguous shadowed vegetation pixels. Hence, our study scope is limited as feature extraction. Besides, there are some relevant studies of the raster-to-vector conversion in the geometry level without the property level (Hori and Tanigawa 1993; Liu and Dori 1999). As for the contour-based vectorization, the main computationally intensive operation is edge detection and polygonalisation, which is in the geometry level (Liu and Dori 1999). The vectorization method by line fitting based on contours and skeletons, proposed by O. Hori *et al.* (1993) is in the geometry level as well. Therefore, the raster-to-vector conversion in this paper only takes the geometry level, instead of the property level, into account.

The proposed image-pyramid-based raster-to-vector conversion (IPBRTVC) framework is not a simple application study, but is a pioneering and innovative study with integration innovation and imitation innovation. The detailed innovation and contributions include the five aspects below:

1. We are the first to propose an IPBRTVC framework for classic object extraction, which can achieve consecutive-scale cartography and vectorization simultaneously.
2. Quality and automation of the raster-to-vector conversion can be ensured by quality control, image segmentation and feature extraction.
3. We firstly propose a method of the consecutive-scale vector presentation, by which problems on arbitrary-scale presentation and aforementioned spatial inconsistency can be solved a lot. In addition, not only can spatial data usage efficiency be improved, but also the data-source error and inconsistency are well avoided.
4. The proposed IPBRTVC framework is more reliable and of higher overall efficiency, because it avoids the correspondence matching, topology errors and additional polyline-to-polygon conversion so that achieves the lower algorithm complexity when compared with the known data processing flows of cartographic generalization.
5. The proposed IPBRTVC framework helps to integrate remote sensing image processing and cartographic generalization and promotes the development of a parallel computing system on spatial vectorization and consecutive-scale presentation, which means the efficiency of multi-scale image processing and cartographic generalization can be greatly increased.

Description of IPBRTVC

Figure 1 shows the IPBRTVC framework, in which details can be optionally modified for optimizing raster-to-vector conversion and consecutive-scale presentation:

First, preprocess the initial image. Second, resample the preprocessed image with a pyramid to achieve the multi-scale representation for a remote sensing image. Third, acquire multi-scale denoised images by quality control. Finally, achieve the multi-scale vectorization presentation with the help of image segmentation and feature extraction, and acquire consecutive-scale vector data by pyramid thought on the basis of the same data source.

According to the framework, Figure 2 illuminates the technique flow, in which parts of the algorithms are substitutable. Here, are the detailed steps:

1. Acquire a multi-scale image via Gaussian pyramid after Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) and band fusion for the remote sensing imagery. Readers can try other pyramids to resample the image, such as Laplacian pyramid and Wavelet pyramid.

1. ArcGIS 10 Help. 2012. <http://resources.arcgis.com/en/help/main/10.1/index.html#/Raster_to_Polygon/0012000000800000/> Accessed 23 January 2019.

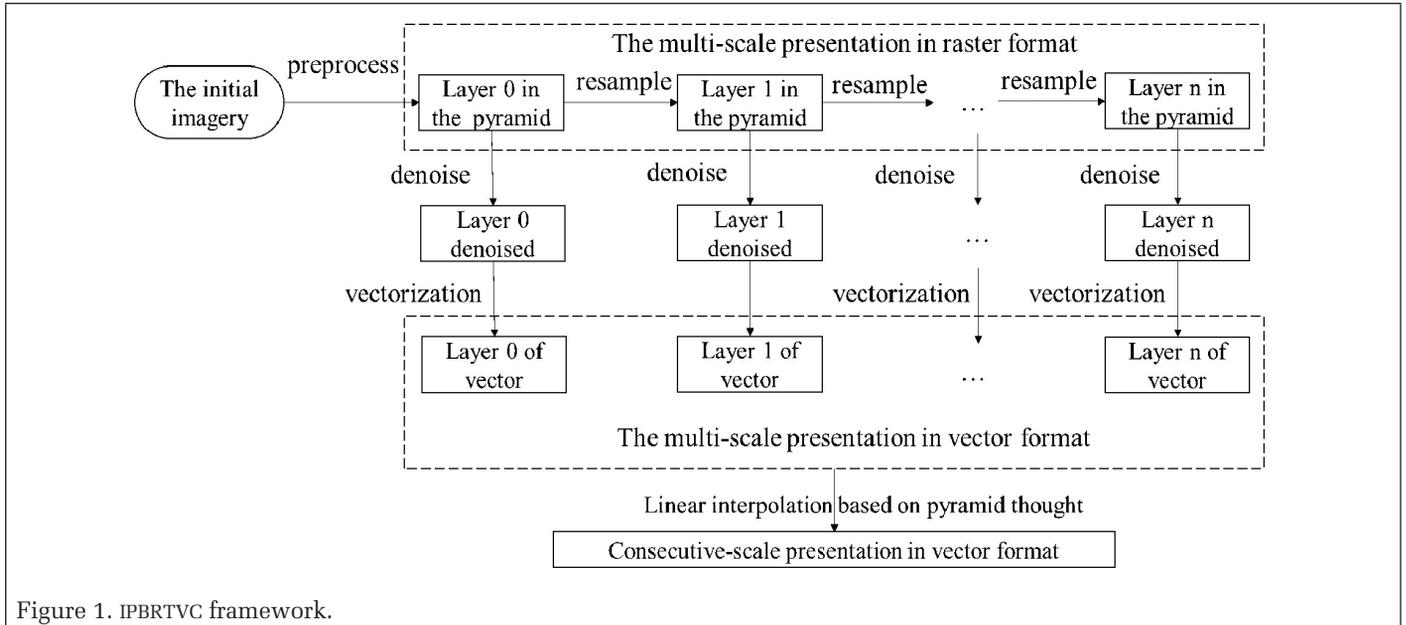


Figure 1. IPBRTVC framework.

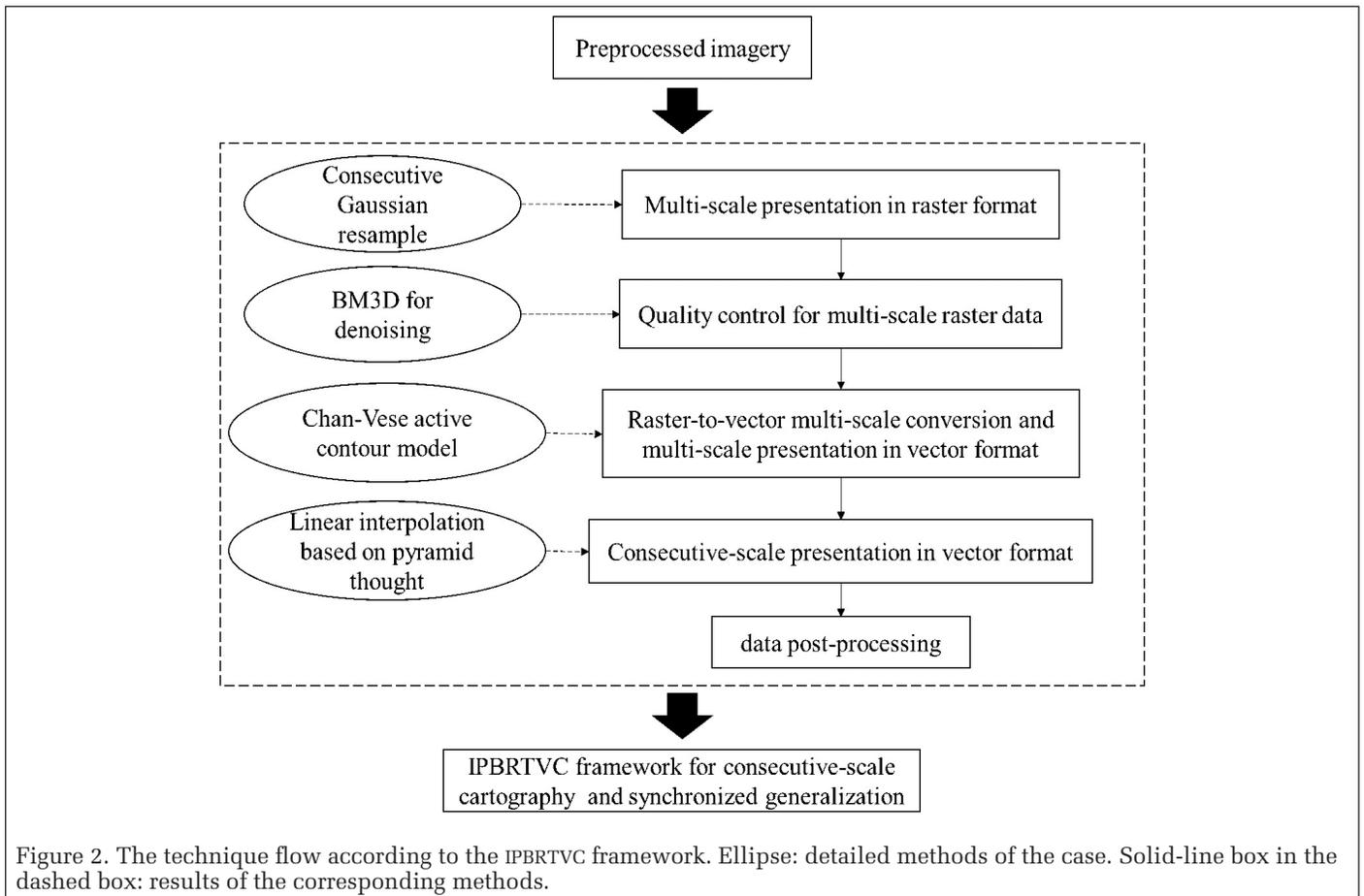


Figure 2. The technique flow according to the IPBRTVC framework. Ellipse: detailed methods of the case. Solid-line box in the dashed box: results of the corresponding methods.

2. The method of Block-Matching and 3D filtering (BM3D) (Dabov *et al.* 2006) is picked in this paper to denoise the multi-scale image for the improvement of image quality and data reliability. Other denoising methods are worth trying as well.
3. Achieve multi-scale automatic vectorization (feature extraction) by the Chan-Vese active contour model. Some other image segmentation methods, such as graph cut, can be applied to raster-to-vector conversion.

What needs noticing is that: (1) what mentioned above is a case and readers can optionally modify some methods and models in the IPBRTVC framework; (2) parallel computing is easily realized by several personal computers (PCs) and graphics processing unit (GPUs) on different layers of a pyramid; and (3) in the future work, the semantics will be considered in the raster-to-vector conversion and consecutive-scale cartography.

In order to highlight the advantages of the proposed framework, the IPBRTVC is compared with others processing flow:

1. The proposed IPBRTVC does not need any operations on corresponding points or objects matching, which helps to avoid matching mistakes and the waste of working time. We compare our work flow with the automated evaluation of building alignments in generalized maps (Zhang *et al.* 2013) (Figure 3a). That method needs alignments matching under different scales and considered matching problems (including many-to-many and partial correspondence in Figure 4) while building the corresponding relationship. However, the proposed IPBRTVC can easily and accurately obtain the coordinate positions of the corresponding points because of the same raster data source and the pyramid method used to obtain the raster-vector multi-scale representation, which means that IPBRTVC does not need any operations on matching of corresponding points or objects (more details can be seen in the Section, Obtaining Consecutive-Scale Vector Data).
2. The proposed IPBRTVC does not need any data simplification (due to detail removal by reducing the spatial resolution) or solve cartographic conflicts (topology errors). We compare our work flow with the one in fully automated generalization of a 1:50 000 map from 1:10 000 data in Figure 3b (Chiang, Leyk, and Knoblock 2011; Jat, Dahiya, and Garg 2014; Stoter *et al.* 2014). Current methods need to simplify the vector data (e.g. thinning the road network) and solve the cartographic conflicts (e.g. eliminating the overlap of polygons or polylines). However, on the basis of the pyramid, the vectorization takes place on the top layer of the pyramid so that the details can be automatically removed and the cartographic conflicts can be avoided. Namely, the IPBRTVC can achieve the results of simplification directly.
3. The proposed IPBRTVC can obtain higher relative accuracy of data processing (i.e. in cartographic generalization, the same-source-based method with the same data processing flow can obtain less errors). Firstly, the IPBRTVC obtains raster-to-vector conversion and multi-scale representation based on the same data source (Figure 1) so there are no data source errors and the integrative processing can reduce the errors produced during the data processing. Secondly, IPBRTVC improves the peak signal-noise ratio (PSNR) by quality control to improve the data processing accuracy. Thirdly, the IPBRTVC realizes the automation of raster-to-vector conversion by Chan-Vese active contour model, which can also improve the conversion accuracy compared with the traditional automated vectorization methods and manual raster-to-vector conversion that depends on the operator's skills and patience. The accumulative errors σ_{overall} during the image processing can be formulated as: $\sigma_{\text{overall}} = \sqrt{\sigma_{\text{source}}^2 + \sigma_{\text{process}}^2}$. Where σ_{source} is the data-source errors and σ_{process} is the data-processing errors. It can be found that relative error of the IPBRTVC is lower than traditional cartographic generalization.
4. The proposed IPBRTVC does not need polyline-to-polygon conversion. Different from the traditional contour-based raster-to-vector conversion (Hori and Tanigawa 1993), IPBRTVC can directly achieve the polygons by using the Chan-Vese active contour model, instead of extracting edges (polylines) first and then obtaining the polygons via polyline-to-polygon conversion. IPBRTVC is of less algorithm complexity.
5. The proposed IPBRTVC contributes to the development of relevant parallel computing. The IPBRTVC realizes the

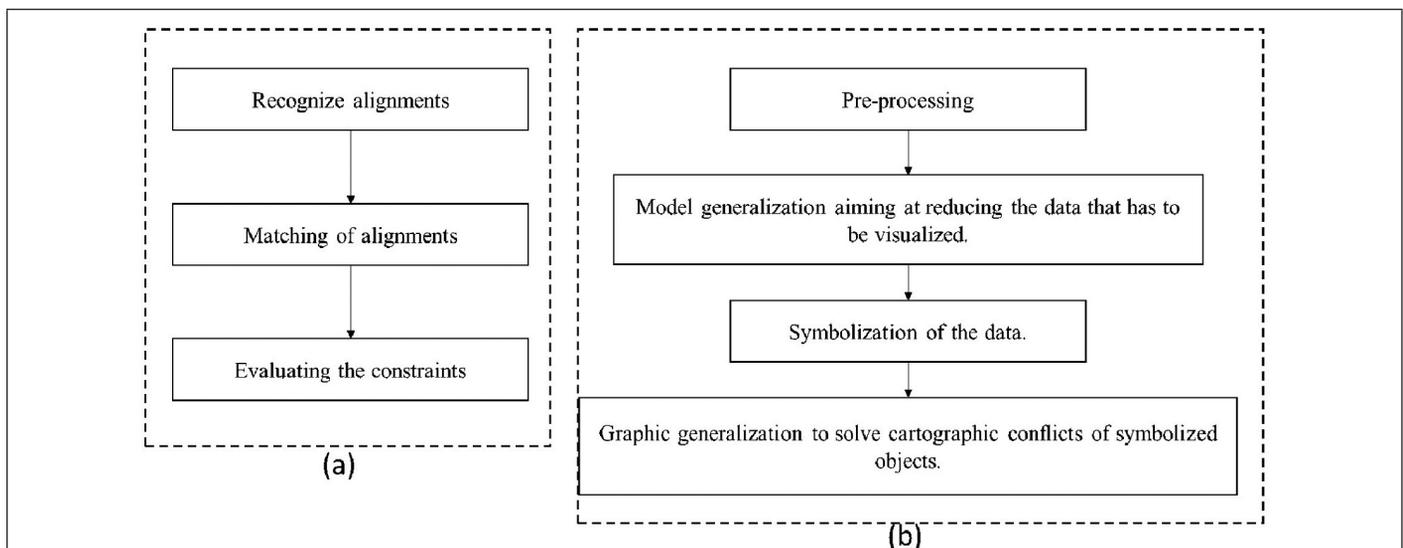


Figure 3. Other cartographic generalization processes. (a) is the work flow in automated evaluation of building alignments in generalized maps (Zhang *et al.* 2013). (b) is the work flow in fully automated generalization of a 1:50 000 map from 1:10 000 data (Stoter *et al.* 2014).

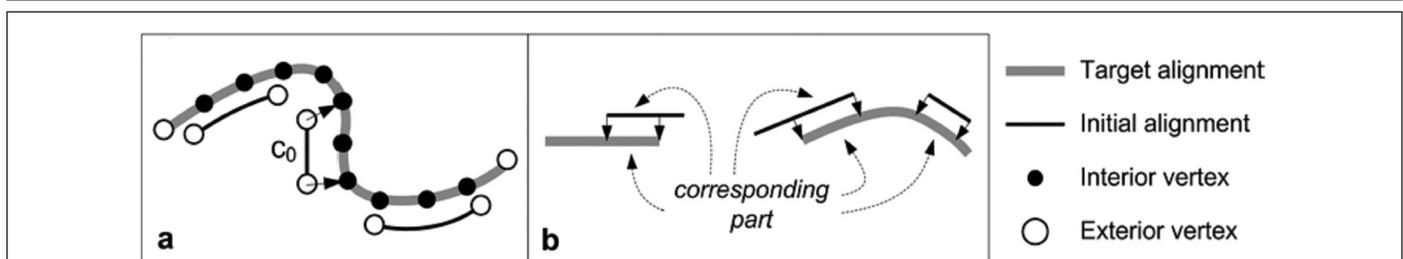
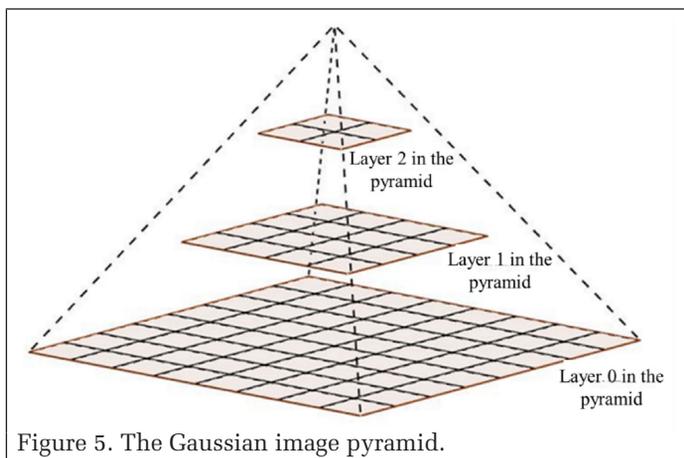


Figure 4. Problems of alignment matching (Zhang *et al.* 2013): (a) many-to-many and (b) partial correspondence.

raster-to-vector conversion and multi-scale representation by pyramid method so that the data on each layer of pyramid can be processed alone. Therefore, it is easy to design and implement relevant parallel computing so that the efficiency of spatial data processing is improved

In all, the proposed IPBRTVC framework is more efficient, more accurate but of the lower algorithm complexity. To avoid ambiguity, we clearly define our study scope as feature extraction of classic object in the geometry level without the property level for consecutive-scale cartography and synchronized generalization.



Gaussian Pyramid

We use a Gaussian method to build a remote sensing image pyramid via Gaussian smooth and resample. Gaussian smooth simulates the process of a retinal image with the help of image blur, which means that the objects are further to eyes, the scales are larger, and the images are more blurred.

A pixel in a resample image layer is a quarter of the corresponding pixel in the last layer (Figure 5). The number of image pyramid layers is concerned with the scale of the initial image and the scale of the top layer required in a certain study. K is the number of pyramid layers, defined by:

$$K = \lceil \log_2(\min(m, n)) \rceil - a \quad (1)$$

where m is the number of rows and n is the number of the columns of the initial image. Parameter a is an arbitrary integer between 0 and $\log_2(\min(m, n))$.

Denoising the Remote Sensing Image

In engineering, image noise is represented by approximate Gaussian noise. We pick the method of BM3D (Dabov *et al.* 2006), the best one to process Gaussian noise at present, from many denoising methods.

Denoising the image by filtering blocks of same colors, BM3D is a corpus of non-local means and wavelet shrinkage. There are two main steps in BM3D, (i.e., initial estimate and final estimate) both of which have the three branch steps of grouping, collaborative filtering and aggregation (Figure 6).

Mean-Squared Error (MSE) and PSNR are used to estimate the effect of BM3D. For the denoised image $I_0(x, y)$ and non-denoised image $I_0(x, y)$, the MSE e_{MSE} and PSNR can be written as:

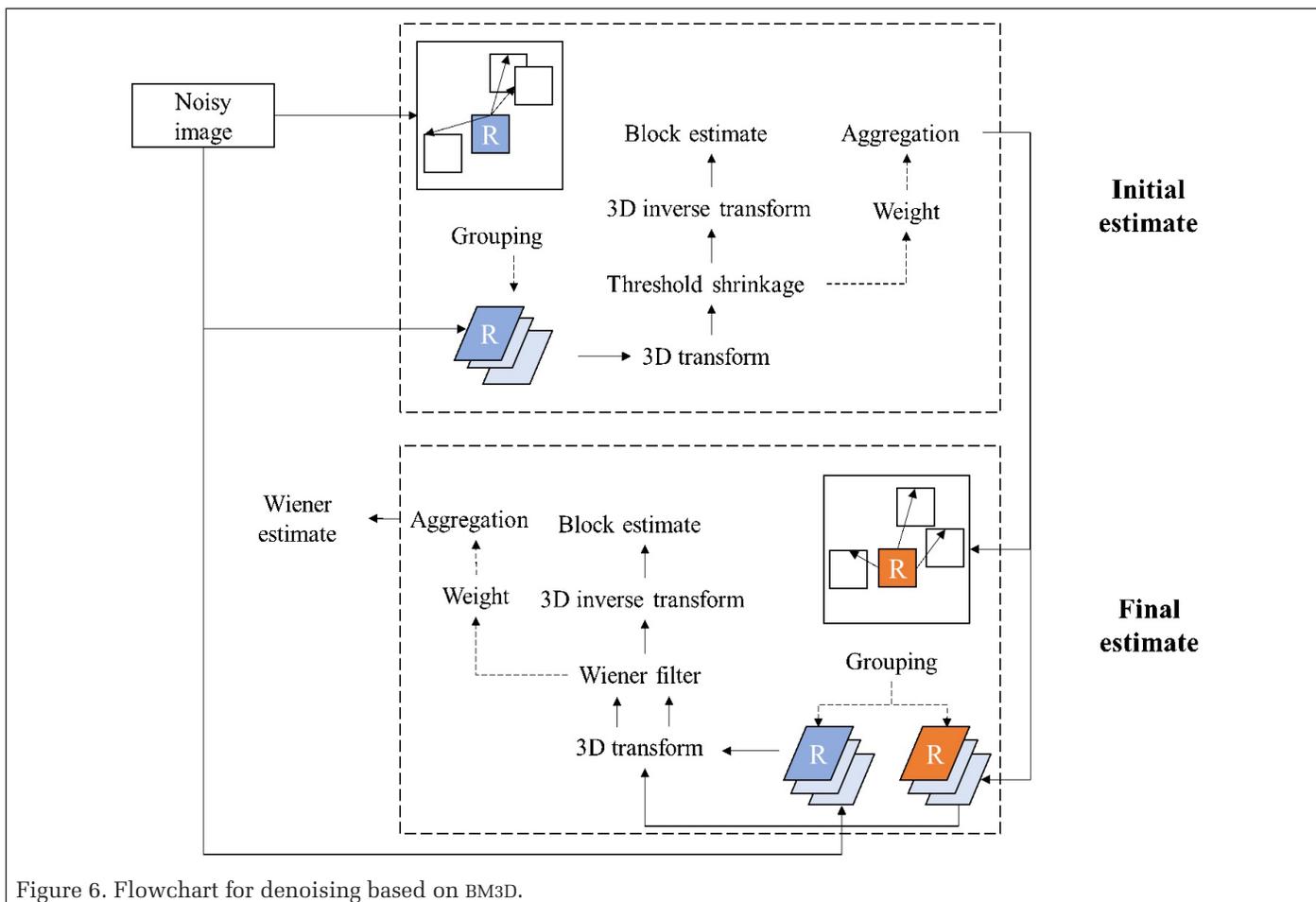


Figure 6. Flowchart for denoising based on BM3D.

$$e_{MSE} = \frac{1}{mn} \sum_{x=1}^m \sum_{y=1}^n (I'_0(x,y) - I_0(x,y))^2 \quad (2)$$

$$PSNR = -10 \lg \frac{e_{MSE}}{s^2} \quad (3)$$

where parameter s is the maximum pixel value of the image. For example, s is 255 in an 8-bit image. PSNR is in decibels (dB), of which a high value means high quality of the image.

Raster-to-Vector Conversion Based on the Chan-Vese Active Contour Model

The Chan-Vese active contour model (Chan and Vese 2001) is modified on the basis of level sets and the Mumford-Shah functional for segmentation, in which traditional parameter-expressed continuous curves are replaced by geometrical ones. When Mumford-Shah segmentation techniques are used, this model has solved the problems that the edges will be smoothed in the classical snakes and active contour models based on the gradient of the image for stopping the process if the image is very noisy. The level set method, instead of the gradient, is used to formulate and solve the minimal partition problem, and then the energy function can be constructed by the pixel gray values so that the minimization problem of the energy function can be considered as the edge detection problem in vectorization.

Assume that the image is $I_0(x,y)$ and C_0 is the boundary of the object to be detected in $I_0(x,y)$. C denotes a close and variable curve. Constants c_1 and c_2 , depending on C , are the averages of $I_0(x,y)$ inside C (or inside(C)) and respectively outside C (or outside(C)). Therefore, the energy function $E(C)$ can be formulated by the inside energy $E_{in}(C)$ and outside energy $E_{out}(C)$ as:

$$e_{MSE} = \int_{inside(C)} |I'_0(x,y) - c_1|^2 dx dy + \int_{outside(C)} |I'_0(x,y) - c_2|^2 dx dy \quad (4)$$

C will coincide with C_0 , the boundary of the object, while $E(C)$ is minimal, i.e., Equation 5:

$$\inf_C \{E_{in}(C) + E_{out}(C)\} \approx 0 \approx E_{in}(C) + E_{out}(C) \quad (5)$$

Obtaining Consecutive-Scale Vector Data

To obtain a consecutive-scale presentation in vector format, we first gain arbitrary-scale vector data by linear interpolation, based on the vector data converted from the multi-scale images generated by a Gaussian pyramid. In a Gaussian pyramid, the width of a pixel in the resampled image is a half of the width of the corresponding pixel in the last layer image. For instance, assume that Layer j ($j = 0, 1, 2, \dots, K$) is between Layer 0 and Layer K (the top layer). Assume further that point (x_0, y_0) is in Layer 0 and its conjugated point in Layer j is (x_j, y_j) . Then, the revelation between the two points is:

$$x_j = 2^{-j} x_0; y_j = 2^{-j} y_0 \quad (6)$$

Define V_m as the vector converted from Layer m in Gaussian image pyramid and V_n as the vector converted from Layer n ($n > m$). (x_m, y_m) and (x_n, y_n) are corresponding points (red points in Figure 7) in V_m and V_n .

Point (x_η, y_η) (the blue point in Figure 7) is in V_η between V_m and V_n and is conjugative with (x_m, y_m) and (x_n, y_n) . Then, we can write the equation:

$$\frac{x_m - x_\eta}{x_\eta - x_n} = \frac{y_m - y_\eta}{y_\eta - y_n} = \frac{\lambda}{1 - \lambda} \quad (7)$$

The equation becomes:

$$x_\eta = (1 - \lambda)x_m + \lambda x_n; y_\eta = (1 - \lambda)y_m + \lambda y_n \quad (8)$$

where constant λ is the scale coefficient. If $\lambda = 0$, $V_\eta = V_m$; if $\lambda = 1$, $V_\eta = V_n$.

Experimental Results

Data Sources and Preprocessing

We conclude this paper by presenting numerical results using our framework on remote sensing imagery, which includes Landsat-8 imagery and DMSP/OLS night-time light imagery. As Figure 8 shows, 8a and 8b are both Landsat-8 images and of 30-meter spatial resolution. The scale of Figure 8a is (in pixels) and 8a is characterized as a simple single water body of linear rivers, while the scale of 8b is (in pixels) and 8b is characterized as a complex water body of linear rivers and polygon lakes. Figure 8c is a DMSP/OLS night-time light image throughout Asia and Europe and has 1-kilometer spatial resolution, reflecting natural contours of urban areas. The scale of Figure 8c is (in pixels). Hence, classic objects in this experiment mainly includes water and urban areas, etc.

Preprocessing includes FLAASH and band fusion in ENVI software. We select band-5 near-infrared (NIR), band-6 short-wave infrared (SWIR1) and band-4 (red) of the Landsat-8 imagery to obtain the false color composite images to distinguish the land and water, which helps to identify water bodies. There is no band combination for the DMSP/OLS night-time light imagery.

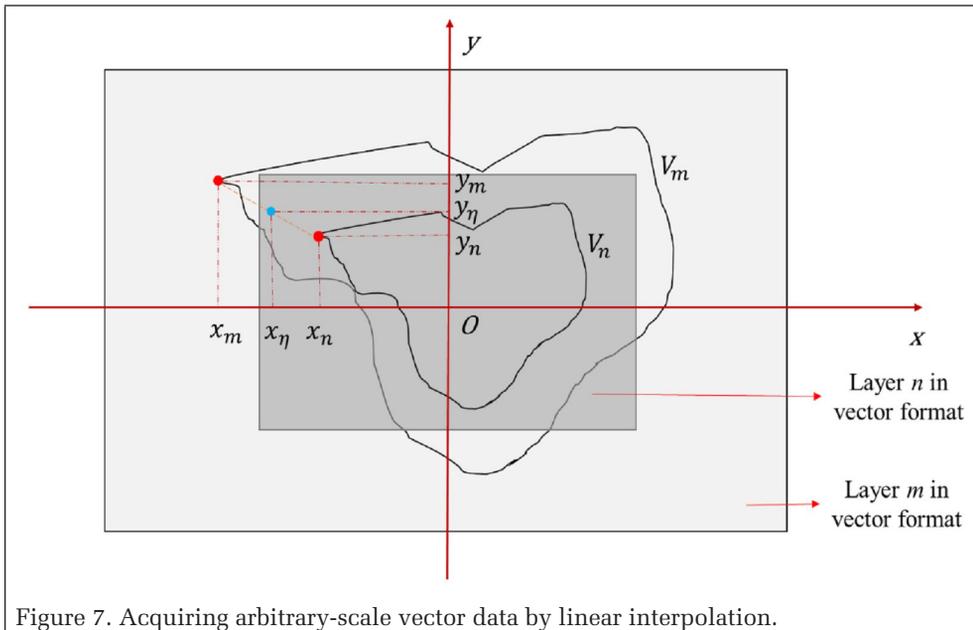


Figure 7. Acquiring arbitrary-scale vector data by linear interpolation.

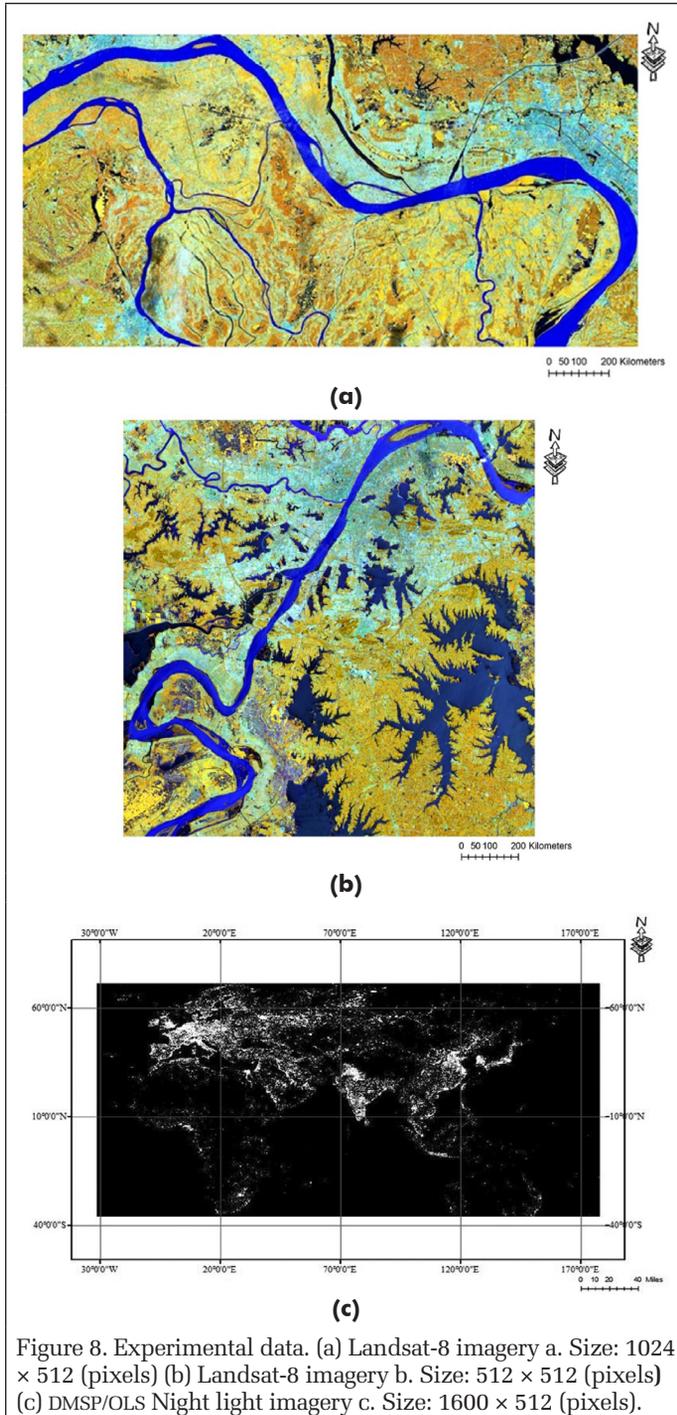


Figure 8. Experimental data. (a) Landsat-8 imagery a. Size: 1024 × 512 (pixels) (b) Landsat-8 imagery b. Size: 512 × 512 (pixels) (c) DMSP/OLS Night light imagery c. Size: 1600 × 512 (pixels).

Results for the Image Pyramid and Denoising

Image Pyramid

Define the preprocessed image as Layer 0 in the image pyramid and define the first resampled image as Layer 1, etc. We conduct a 4-layer pyramid for each group of imagery (Figure 9). We can find that the spatial resolution and information of the imagery decreases as the number of resamples increases for one group of imagery.

Denoising

After Gaussian pyramid transformation, the signal-noise ratio (SNR) and ability to detect objects decrease. So, it is necessary to denoise the multi-scale images to guarantee the quality of the vectorization later. We denoise the images in MATLAB

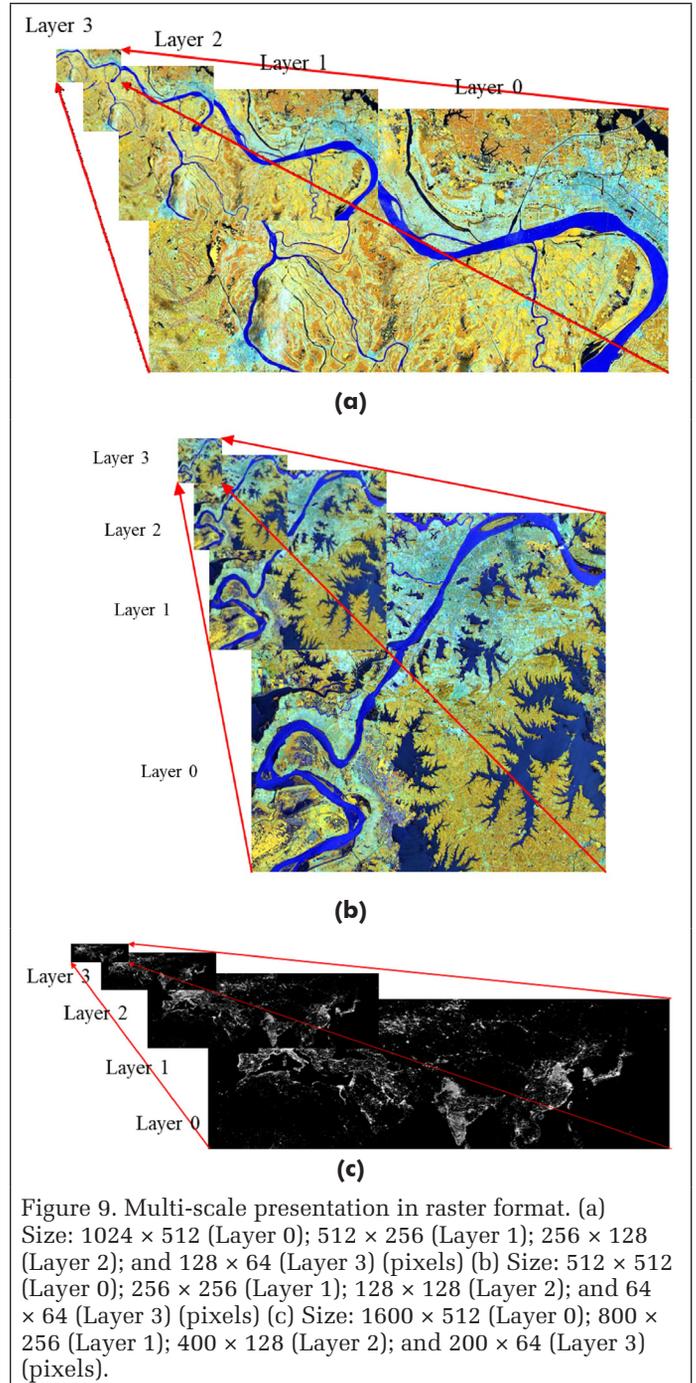


Figure 9. Multi-scale presentation in raster format. (a) Size: 1024 × 512 (Layer 0); 512 × 256 (Layer 1); 256 × 128 (Layer 2); and 128 × 64 (Layer 3) (pixels) (b) Size: 512 × 512 (Layer 0); 256 × 256 (Layer 1); 128 × 128 (Layer 2); and 64 × 64 (Layer 3) (pixels) (c) Size: 1600 × 512 (Layer 0); 800 × 256 (Layer 1); 400 × 128 (Layer 2); and 200 × 64 (Layer 3) (pixels).

software using the BM3D method. We set the average value of Gaussian noise as 0 and the variance as 0.05. The parameter configurations are given in Table 1.

We evaluate the multi-scale denoising effect by PSNR and the time of processing. Generally, the imagery is artificially identifiable if its PSNR surpasses 30 decibels. If PSNR is higher, then the denoising effect is better and the imagery has higher quality. The results of multi-scale denoising are given in Table 2.

It is easy to find the large difference between before-denoising and after-denoising from Table 2, which generally performs the improved PSNR and quality of the imagery after denoising. The time of processing varies as well, due to the different information richness and scales of images in different spatial resolutions. Generally, larger-scale images have

Table 1. Parameter configurations for BM3D.

Parameters	Parameter Configurations
Initial estimation	
Size of the image block	8×8 (in pixel)
Searching region	39×39 (in pixel)
Number of blocks most similar to the reference block	16
Two-dimensional transform method	Wavelet transform
One-dimensional transform method	Hadamard transform
Hard threshold	2.7
Final estimation	
Size of the image block	8×8 (in pixel)
Searching region	39×39 (in pixel)
Number of blocks most similar to the reference block	32
Two-dimensional transform method	Wavelet transform
One-dimensional transform method	Hadamard transform

Table 2. PSNR and processing time of denoising multi-scale images.

Imagery	Layers	Size (in Pixel)	PSNR (in dB)		Time (Seconds)
			Before denoising	After denoising	
Landsat-8 imagery a	Layer 0	1024 × 512	14.1514	61.4893	1385.9882
	Layer 1	512 × 256	14.0085	61.7071	254.8328
	Layer 2	256 × 128	13.9618	61.5523	65.8086
	Layer 3	128 × 64	13.8876	61.3551	13.8375
Landsat-8 imagery b	Layer 0	512 × 512	13.9272	61.5082	515.4219
	Layer 1	256 × 256	13.7563	61.5868	127.7306
	Layer 2	128 × 128	13.6744	61.4169	32.1745
	Layer 3	64 × 64	13.5866	61.0990	6.9037
DMSP/OLS night light imagery c	Layer 0	1600 × 512	15.6689	67.2347	1751.8115
	Layer 1	800 × 256	15.5803	67.8878	619.3316
	Layer 2	400 × 128	15.5355	67.7537	145.1896
	Layer 3	200 × 64	15.5379	67.0483	34.2747

Table 3. Parameter configurations for vectorization.

Imagery	Layers	Size (in Pixel)	Iteration Times	Size of the Vectorizing Window
Landsat-8 imagery a	Layer 0	1024 × 512	3000	300 × 300
	Layer 1	512 × 256	2500	150 × 150
	Layer 2	256 × 128	2500	100 × 100
	Layer 3	128 × 64	1500	30 × 30
Landsat-8 imagery b	Layer 0	512 × 512	10 000	300 × 300
	Layer 1	256 × 256	5000	100 × 100
	Layer 2	128 × 128	2000	50 × 50
	Layer 3	64 × 64	1000	15 × 15
DMSP/OLS night light imagery c	Layer 0	1600 × 512	8000	300 × 300
	Layer 1	800 × 256	6000	90 × 90
	Layer 2	400 × 128	5000	50 × 50
	Layer 3	200 × 64	2000	30 × 30

more complex object structures so that the time complexity of denoising is higher. Taking Layer 0 of Figure 9c and Layer 3 of 9b for example, the denoising time of the former is 1751.8115 seconds, which is relatively longer because of its large scale (1600 × 512 dB) and thin complex boundaries, while the denoising time of the latter is only 6.9037 seconds, which is the result of the smaller-scale (64 × 64 dB) image with less information.

Vectorization and Multi-Scale Representation in Vector Format

We select the Chan-Vese active contour method to achieve the vectorization in MATLAB software. We pick relevant parameters and methods according to objects that remain to be vectorized. The iteration times and the size of the vectorizing window are the main parameters to be considered. We achieve the automatic vectorization of simple single objects. The specific parameters are given in Table 3.

Scales of the vector data, which is converted from multi-scale raster data, correspond with the scales of multi-scale raster. In Figure 10, red boundaries are the boundaries of water bodies after vectorization in Figure 10a and 10b, and boundaries of urban areas in 10c. As the scale of vectorization changes, the differences in the importance of objects can be presented via simplifying objects. Smaller-scale vector data has richer information.

Obtaining Consecutive-Scale Vector Data

We have two cases where $\lambda = 0.3$ and $\lambda = 0.7$ (Equation 8) based on the vector data converted from Layer 0 and Layer 1 of the Figure 9b pyramid.

As for the vector converted from Layer 0, the vector data converted from Layer 1 simplifies the information, and the information richness of the vector that is obtained via linear interpolation is in between. The scale coefficient λ decides the simplifying level of the interpolation vector relative to the vector converted from Layer 0. The simplifying level is higher while λ is bigger ($\lambda \in [0,1]$), which means the interpolation vector is more similar to the vector converted from the last layer pyramid.

Concluding Remarks and Discussions

In this paper, we propose an image-pyramid-based raster-to-vector conversion (IPBRTVC) framework for consecutive-scale cartography and synchronized generalization based on the data source consistency. The proposed IPBRTVC framework is a grand design of spatial data processing flows (i.e. integrating remote sensing and GIS), and a pioneering and innovative study with integration innovation and imitation innovation which are both included in the independent innovation. The IPBRTVC can integrate any state-of-the-art algorithms of image processing and image classification into our framework, which is more contributing than optimizing one algorithm of cartographic generalization and helps to realize the great innovation of spatial data processing with the higher efficiency. The experimental results verify the feasibility, reliability and generalizability of IPBRTVC. The main conclusions are as follows:

1. Tiny objects can be simplified in the large-scale imagery by decreasing the spatial resolution so that the importance of large objects is emphasized. We can obviously learn the importance of different objects with multi-scale observation, which helps to comprehensively understand the geographical phenomena and process and contributes to model building and optimizing in remote sensing inversion.
2. Quality control before vectorization can substantially increase the PSNR of the imagery and improve the image quality to a great extent.
3. It proves possible to achieve the automation of raster-to-vector conversion by using the energy function in the

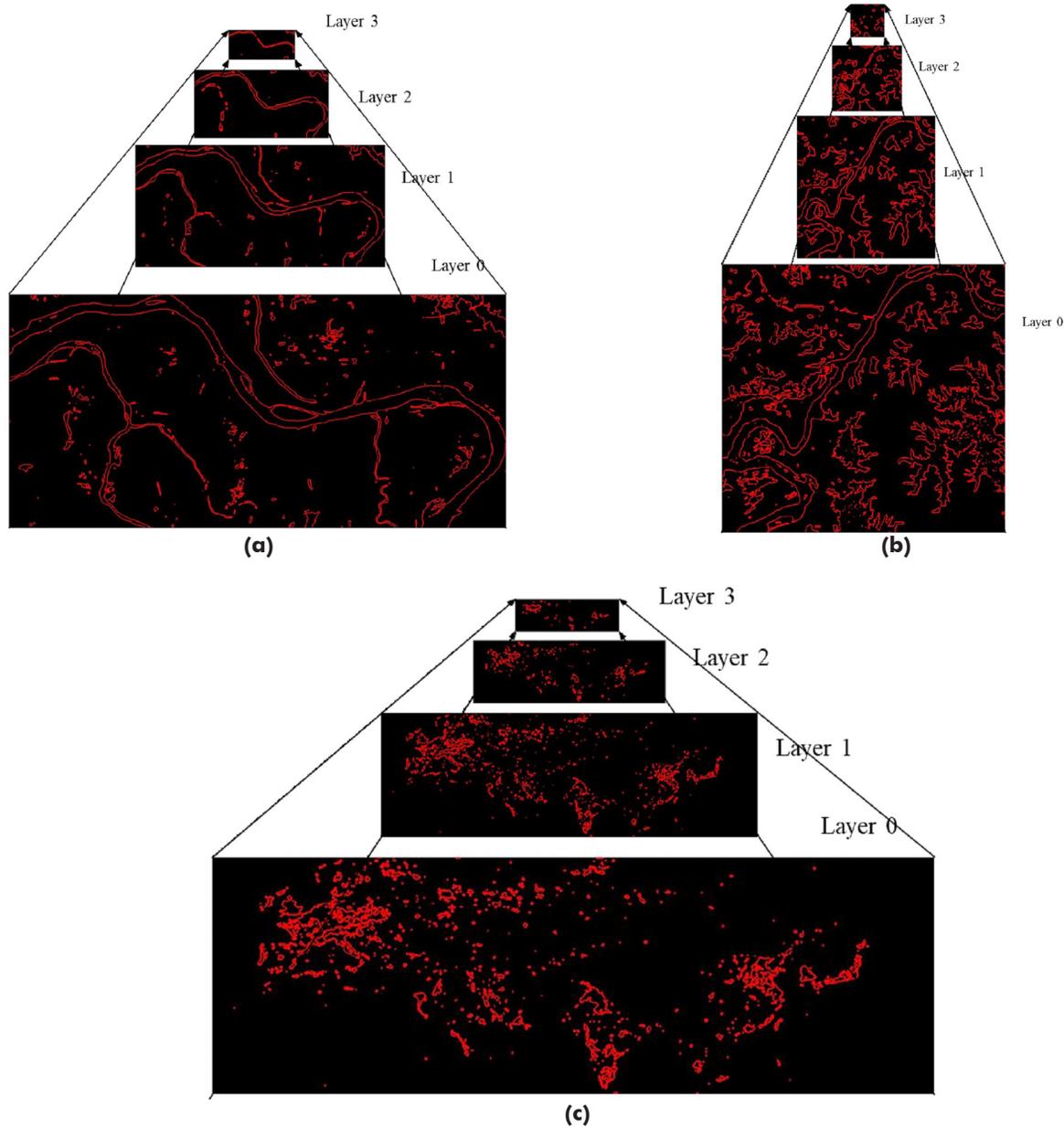


Figure 10. Multi-scale presentation in vector format. The red is vectorized boundaries. (a) Size: 1024×512 (Layer 0); 512×256 (Layer 1); 256×128 (Layer 2); and 128×64 (Layer 3) (pixels) (b) Size: 512×512 (Layer 0); 256×256 (Layer 1); 128×128 (Layer 2); and 64×64 (Layer 3) (pixels) (c) Size: 1600×512 (Layer 0); 800×256 (Layer 1); 400×128 (Layer 2); 200×64 (Layer 3) (pixels).

Chan-Vese active contour model to search and detect object boundaries under the strict quality control.

4. The pyramid method can promote raster-to-vector multi-scale conversion and multi-scale presentation in raster and vector formats and helps to acquire arbitrary-scale vector data.
5. Compared with the traditional contour-based raster-to-vector conversion, the IPBRTVC framework achieves the higher data quality and overall processing efficiency, and is of the lower algorithm complexity, because IPBRTVC avoids matching problems, cartographic conflicts (topology errors), and the additional polyline-to-polygon conversion due to the pyramid-based method and Chan-Vese active contour model with quality control.
6. This framework may change traditional geographic processing methods, which integrates remote sensing data processes and cartographic data processes, contributes to

the development of relevant parallel computing system, improves the efficiency of spatial data process and renovates the procedure of data process in remote sensing and cartography, and it may even promote the approach in the integration of remote sensing and GIS.

The IPBRTVC framework is a preliminary exploratory study of consecutive-scale cartography and synchronized generalization in the geometry level without the property level so that there are still some problems that need to be studied deeply in the future work, that is to say that we will study semantic problem (i.e. spatial information, attribute information and temporal information) in the raster-to-vector conversion and consecutive-scale cartography. The future work also includes: (1) extracting other classic objects in high-resolution image; and (2) designing a relevant parallel computing system to achieve high efficiency of consecutive-scale cartography and synchronized generalization based on massive remote sensing images.

Acknowledgments

The authors thank the National Natural Science Foundation of China (NSFC) (Grant Nos. 41771493 and 41101407), the Natural Science Foundation of Hubei Province (Grant Nos. 2014CFB377, and 2010CDZ005), China, the Wuhan Youth Science and technology plan (Grant Nos. 2016070204010137) and the self-determined research funds of CCNU from the basic research and operation of MOE (Grant No. CCNU15A02001) for supporting this work. The authors are grateful for the comments and contributions of Editor-in-Chief: Prof. Dr. Alper Yilmaz, the anonymous reviewers and the members of the editorial team.

References

- Adelson, E. H., C. H. Anderson, J. R. Bergen, P. J. Burt and J. M. Ogden. 1984. Pyramid methods in image processing. *RCA Engineer*.
- Ameri, F., M.J.V. Zoj and M. Mokhtarzade. 2016. Multi-criteria, graph-based road centerline vectorization using ordered weighted averaging operators. *Photogrammetric Engineering & Remote Sensing* 82 (2):107–120.
- Brewer, C. A., B. P. Buttenfield and E. L. Usery. 2009. Evaluating generalizations of hydrography in differing terrains for the National Map of the United States. In *Proceedings of the International Cartographic Conference*.
- Buttenfield, B. P., L. V. Stanislawski and C. A. Brewer. 2010. Multiscale representations of water: Tailoring Generalization Sequences to Specific Physiographic Regimes. In *Proceedings of the Giscience*.
- Chan, T. F. and L. A. Vese. 2001. *Active Contours without Edges*. IEEE Press.
- Chiang, Y. Y., S. Leyk and C. A. Knoblock. 2011. Efficient and robust graphics recognition from historical maps. In *Proceedings of the International Conference on Graphics Recognition: New Trends and Challenges*.
- Dabov, K., A. Foi, V. Katkovnik and K. O. Egiazarian. 2006. Image denoising with block-matching and 3D filtering. *Proceedings of SPIE*: 606414.
- Greenberg, J. A., S. Z. Dobrowski and S. L. Ustin. 2005. Shadow allometry: Estimating tree structural parameters using hyperspatial image analysis. *Remote Sensing of Environment* 97 (1):15–25.
- Greenspan, H., S. Belongie, R. Goodman, P. Perona, S. Rakshit and C. H. Anderson. 2002. Overcomplete steerable pyramid filters and rotation invariance. Paper read at Computer Vision and Pattern Recognition. 1994. In *Proceedings of the 1994 IEEE Computer Society Conference on CVPR '94*.
- Hori, O. and S. Tanigawa. 1993. Raster-to-vector conversion by line fitting based on contours and skeletons. In *Proceedings of the International Conference on Document Analysis and Recognition*.
- Jat, M. K., S. Dahiya and P. K. Garg. 2014. Building extraction from high resolution satellite images. In *Proceedings of the 2013 IEEE 3rd International Advance Computing Conference (IACC)*.
- Jimenez, J. and J. L. Navalon. 1982. Some experiments in image vectorization. *IBM Journal of Research & Development* 26 (6):724–734.
- Kim, M., T. A. Warner, M. Madden and D. S. Atkinson. 2011. Multi-scale GEOBIA with very high spatial resolution digital aerial imagery: Scale, texture and image objects. *International Journal of Remote Sensing* 32 (10):2825–2850.
- Li, Z. and Y. H. Choi. 2002. Topographic map generalization: Association of road elimination with thematic attributes. *Cartographic Journal* 39 (2):153–166.
- Li, Z. and Q. Zhou. 2012. Integration of linear and areal hierarchies for continuous multi-scale representation of road networks. *International Journal of Geographical Information Science* 26 (5):855–880.
- Liu, W. and D. Dori. 1999. From raster to vectors: Extracting visual information from line drawings. *Pattern Analysis & Applications* 2 (1):10–21.
- Rosenfeld, A. and L. S. Davis. 1976. A note on thinning. *IEEE Transactions on Systems Man & Cybernetics smc-6* (3):226–228.
- Saeed, K., M. Rybnik and M. Tabedzki. 2001. Implementation and advanced results on the non-interrupted skeletonization algorithm. *Computer Analysis of Images and Patterns*: 601–609.
- Saeed, K., M. Tabedzki, M. Rybnik and M. Adamski. 2010. K3M: A universal algorithm for image skeletonization and a review of thinning techniques. *International Journal of Applied Mathematics and Computer Science* 20 (2):317–335.
- Simoncelli, E. P. and W. T. Freeman. 1995. The steerable pyramid: a flexible architecture for multi-scale derivative computation. In *Proceedings of the International Conference on Image Processing*.
- Stoter, J., M. Post, V. V. Altena, R. Nijhuis and B. Bruns. 2014. Fully automated generalization of a 1:50k map from 1:10k data. *Cartography & Geographic Information Science* 41 (1):1–13.
- Weibel, R. 1997. Generalization of spatial data: Principles and selected algorithms. 99–152.
- Woodcock, C. E. and A. H. Strahler. 1987. The factor of scale in remote sensing. *Remote Sensing of Environment* 21 (3):311–332.
- Wu, H. and Z. Li. 2009. Scale issues in remote sensing: A review on analysis, processing and modeling. *Sensors* 9 (3):1768–1793.
- Yang Wei, A.T.-H. 2016. Road centerline extraction from crowdsourcing trajectory data. *Geography and Geo-Information Science* 32 (3):1–7.
- Zhang, X., J. Stoter, T. Ai, M. J. Kraak and M. Molenaar. 2013. Automated evaluation of building alignments in generalized maps. *International Journal of Geographical Information Science Ijgis* 27 (8):1550–1571.
- Zhao, Y. and H. Bo. 2013. Multi-temporal MODIS and Landsat reflectance fusion method based on super-resolution reconstruction. *Journal of Remote Sensing* 17 (3):590–608.

Land Cover Classification in Combined Elevation and Optical Images Supported by OSM Data, Mixed-level Features, and Non-local Optimization Algorithms

Dimitri Bulatov, Gisela Häufel, Lukas Lucks, and Melanie Pohl

Abstract

Land cover classification from airborne data is considered a challenging task in Remote Sensing. Even in the case of available elevation data, shadows and strong intra-class variations of appearances are abundant in urban terrain. In this paper, we propose an approach for supervised land cover classification that has three main contributions. Firstly, for the cumbersome task of training data sampling we propose an algorithm which combines the freely available OpenStreetMap data with the actual sensor data and requires only a minimum of user interaction. The key idea of this algorithm is to rasterize the vector data using a fast segmentation result. Secondly, pixel-wise classification may take long and be quite sensitive to the resolution and quality of input data. Therefore, superpixel decomposition of images, supported by a general framework on operations with superpixels, guarantees fast grouping of pixel-wise features and their assignment to one of four important classes (building, tree, grass and road). Particularly for extraction of street canyons lying in the shadowy regions, high-level features based on stripes are introduced. Finally, the output of a probabilistic learning algorithm can be post-processed by a non-local optimization module operating on Markov Random Fields, thus allowing to correct noisy results using a smoothness prior. Extensive tests on three datasets of quite different nature have been performed with two probabilistic learners: The well-known Random Forest and by far less known Import Vector Machine are explored. Thus, this work provides insights about promising feature sets for both classifiers. The quantitative results for the ISPRS benchmark dataset Vaihingen are promising, achieving up to 94.5% and 87.1% accuracy on superpixel and on pixel level, respectively, despite the fact that only around 10% of available labeled data were used. At the same time, the results for two additional datasets, validated with the autonomously acquired training data, yielded a significantly lower number of misclassified superpixels. This confirms that the proposed algorithm on training data extraction works quite well in reducing errors of second kind. However, it tends to extract predominantly huge and easy-to-classify areas, while in complicated, ambiguous regions, first type errors often occur. For this and other algorithm shortcomings, directions of future research are outlined.

Introduction

Motivation

Land cover classification, especially in urban and semi-urban environment, is a key step for creating semantic 3D models

Fraunhofer Institute of Optronics, System Technologies and Image Exploitation Gutleuthausstr. 1, 76275 Ettlingen, Germany (dimitri.bulatov@iosb.fraunhofer.de).

from airborne sensor data. As mentioned e.g. by Bulatov et al. (2014), the advantages of semantic models are: Higher reality content, flexible level of compression, as well as better interpretability and interoperability on the non-expert users' part. Buildings are modeled on a desired level of detail, trees and vehicles are placed at positions they have been detected and are represented by geo-specific models from a library, etc. The emphasis of that and comparable studies (Haala, 2005; Lafarge and Mallet, 2012) was predominantly laid on reconstruction of complicated objects, in particular, building roof types. However, not much effort was invested into a precise and reliable subdivision of the underlying terrain into different classes. Therefore, at most, few very discriminative features, such as elevation and color indexes (like NDVI, normalized difference vegetation index), were considered to separate buildings from trees, roads from vehicles, water bodies from grass areas.

In order to perform a more systematic preparation of data for the aforementioned reconstruction task, classification of the underlying terrain is necessary. In real-case scenarios, there are plenty of factors hindering a correct assignment of pixels to classes, which would later result in incorrect building outlines, wriggled street courses, etc. Examples of assignment errors are sometimes related to seldom or overlapping classes, such as hills covered by shrubbery and grass, destroyed buildings, bridges in a non-negligible height over the ground, etc. Even without taking these anomalies into account or setting them right at a later point, latest developments brought about an extremely broad spectrum of airborne sensors and their products. Taking into account rather heterogeneous scenes to be captured, this may provide very particular patterns of texture, distributions of shadows and types of objects to be classified. Under these circumstances, it is not realistic to obtain a good classification result without (1) taking into account examples of the data currently investigated, (2) computing sometimes sophisticated features, and (3) applying a classification algorithm supposed to learn thresholds on features for separation of training data and thus classify the test data. In other words, classification approaches have three major ingredients: Training data, feature set, and learning algorithm.

Contributions

In this work, we will focus on the three components mentioned above. Firstly, it is interesting to investigate to what

Photogrammetric Engineering & Remote Sensing
Vol. 85, No. 3, March 2019, pp. 179–195.
0099-1112/18/179–195

© 2019 American Society for Photogrammetry
and Remote Sensing
doi: 10.14358/PERS.85.3.179

extent freely available sources are useful for automatic acquisition and preparation of training examples, especially while being meaningfully combined with actual sensor data. One of the main trends in today's way of collecting and processing geographical information is the shift from the classical Remote Sensing, equipped with expensive sensors, to a more crowd-sourced imagery (Rumpler *et al.*, 2012). To achieve this, the worldwide online community OpenStreetMap (OSM) contains freely available vector data for huge areas of the world. However, two principal problems are that it is usually given as vector and not as raster data, and that it can be obsolete, irrelevant and incorrect. In order to cope with these problems and to be able to use this open source data for classification, its fusion with most discriminative features from the sensor data is employed within a statistical approach.

Secondly, we will analyze which feature sets are hypothetically conceivable for classification rather than necessity of individual features. Different hand-crafted feature sets are being used in Computer Vision and Remote Sensing, such as texture features, histograms of oriented gradients, dense descriptors, morphological profiles, and many others. We developed an approach for segment-wise feature collection, for which theoretically any set of features can be applied. Additionally to color and elevation values, we utilize in this work morphological profiles and responses of a rotationally-invariant filter bank. To this standard set of features, generalized canyon-like stripes have been added which represent an example of high-level features applied superpixel-wise. Doing so, we wish to generalize the work of Warnke (2017), Bulatov and Warnke (2017), who performed feature selection for extraction of road class using unfiltered image channels, filter banks, and road-like stripes. The conclusion could be drawn that the stripe-based features were declared as quite important in all scenarios. This motivated us to explore these features for multi-class problems as well as to consider building-like and road-like features based on stripes. Thus, there will be four classes (*building, road, tree, grass*), which are among the most popular in remote sensing and which related contributions are mainly dealing with. Another difference is that there will be not single, but groups of feature sets for which we perform evaluation.

The third aspect concerns the choice of learning algorithm and postprocessing. It is well-known that performance of a classifier is highly sensitive to the chosen set of features. On the one hand, a classifier with a property to be robust to many redundant and irrelevant features is the state-of-the-art Random Forest (RF), (Breiman, 2001). On the other hand, we considered the Import Vector Machine (IVM) classifier due to Zhu and Hastie (2005.) Because of its very elegant mathematical formulation, we decided to provide a short description of this classifier and to test its performance. Another reason for choosing RF and IVM is that both aforementioned classifiers are probabilistic. This means that the results of learning algorithm reflect the likelihood of a pixel (or, in our case, a superpixel) to belong to a certain class: While in Random Forest, the posterior probabilities are derived from the relative number of trees having a certain class as output, it was shown by Roscher *et al.* (2012) that also IVM provides posterior probabilities. As for postprocessing, the authors were inspired by the contributions of Schindler (2012), Wegner *et al.* (2015), Montoya-Zegarra *et al.* (2015), Bulatov and Warnke (2017), Sun *et al.* (2017). They have in common that the classification task was embedded into an optimization process on Markov or Conditional Random Fields (CRFs) with (negative) classification output as the data term. Apart from the usual smoothness priors, namely, encouraging neighboring pixels or superpixels to belong to the same class, Wegner *et al.* (2015) also established priors exploiting connectivity property of

road networks in order to perform road pixel extraction. For this purpose, minimum cost paths containing hypothesized road superpixels were incorporated. Additionally, Montoya-Zegarra *et al.* (2015) generalized the priors for the building class using 3D alpha-shapes (Edelsbrunner and Mücke, 1994). In both cases, the optimization was based on robust potentials involving higher-order cliques. However, because of the exploding number of parameters, such as the alpha value for alpha-shapes and different smoothness parameters, it is difficult to track their influence on the results. Estimation of these parameters using training data is possible, as e.g., Sun *et al.* (2017) mentioned, but cumbersome. In addition, any non-local optimization algorithm is strongly correlated with the data term, which is in the focus of our work. For these two reasons, we will merely consider a symmetric smoothness prior.

To summarize our contributions, we will provide:

1. a description how the OSM vector data is converted to training data by means of a statistical approach,
2. a generalization of high-level features based on stripes, whereby the aspect of efficient feature collection over stripes, superpixels, and other kinds of segments will be elucidated,
3. and, finally, an accuracy assessment of two classifiers (in particular IVM) and superpixel-wise non-local optimization.

Organization

The paper is structured as follows: In the next section, we will review the related publications on training data extraction and classification of important land cover classes. Then, the preliminaries describing the provided input, the desired output, and the essential preprocessing steps are presented. Our contributions to classification will be in the afterwards followed by description of the employed datasets and the main insights of their evaluation. Finally, the main conclusions and directions of future research will be given.

Related Works

The task of image classification is to generate land cover maps with building, road, vegetation, and other relevant classes. Because of the diversity of appearances of these classes, searching training examples for each class may be a tedious procedure. Even though excellent surveys on crowd-sourcing geographical data exist (Heipke, 2010), in this section, we will describe the main sources freely available data can be obtained from. In the second part, ways to integrate this data into land cover classification algorithms as well as main tools employed for classification with (Lopes *et al.*, 2017; Yuan and Cheriyyadat, 2013; Mills *et al.*, 2015; Mattyus *et al.*, 2015; Huang *et al.*, 2015; Kaiser *et al.*, 2017; Huang *et al.*, 2015) and without OSM data (Shackelford and Davis, 2003a and b; Kluckner *et al.*, 2009; Bellens *et al.*, 2008; Maas *et al.*, 2017; Fröhlich *et al.*, 2013; Gerke and Xiao, 2014; Guo *et al.*, 2011; Huang *et al.*; 2014, Huang and Zhang, 2013; Huang *et al.*, 2017, Sun *et al.*; 2017, Zhang *et al.*, 2015) will be presented.

Overview of Sources

In a few seldom cases, such of the German federal states Thuringia¹, North Rhine-Westphalia, and Berlin, widespread, accurate and free GIS-data are offered. They contain ortho-photos with a resolution of 20 cm, elevation data, but also additional land use information (authoritative real estate cadaster information system, ALKIS), where road-use areas are defined by polygons, and 3D building models are available

1. <http://www.geoportal-th.de/de-de/downloadbereiche/downloadoffenegeodatenth%C3%BCrtingen.aspx>

as LOD1 or LOD2 representation. At a more coarse resolution, other regions of the world follow the example and declassify the information about land cover, such as it was done in Bolivia.² A classification algorithm is then supposed to refine these results, or the data can be used for classification of similar data material, like in neighboring countries. For this latter task, transfer-learning methods, like the one of Vogt *et al.* (2017), can be applied. Furthermore, sporadically land cover classification results can be encountered in crowd-sourced platforms, like Geo-Wiki³ described by Fritz *et al.* (2009) or WUDAPT (World Urban Database and Access Portal Tools; Mills *et al.*, 2015). In the first contribution due to Fritz *et al.* (2009), users are encouraged to submit and to validate classification results and datasets at different resolutions, whereby closer-range data, fragments of Google Maps and Google Earth screen-shots are rapidly gaining popularity. Beside the non-ubiquity of the data, Fritz *et al.* (2009) point out that one serious problem is credibility, because non-cooperative participants may wish to abuse the tool voluntarily. The idea behind the second contribution, WUDAPT is to provide acquisition, dissemination, and storage of climate data in urban terrain. Therefore, materials available in urban infrastructure (building density, building forms) are additionally taken into account and stored as raster data. The OSM shapefiles are thereby used for verification and updating the LCZ (Local Climate Zones) data.

However, in wider areas of the world only vector data, provided by OpenStreetMap, are available. For this data, studies about correctness and completeness have been accomplished, e.g. Haklay (2010) and thus it will be also our default tool for extracting training data for land cover classification. In the next section, we will review the related work combining information from OSM or comparable data and actual sensor data.

Classification Supported by Freely Available Data

One problem about OSM data is that it is given in vector format, that is, as points (positions of tree trunks), closed polygonal chains (buildings and meadows) as well as open polygonal chains (roads). Concerning detection of individual trees using the seed points given by OSM geo-referenced coordinates of tree trunks, no previous works are known to the authors. Most likely, the approach based on snakes (Butenuth and Heipke, 2012) could be applied since the main input necessary for it, namely seed points and approximate extensions are available. However, it is cheaper to select huge forest regions manually.

Though it is not a big deal to rasterize building polygons (Kaiser *et al.*, 2017, Rumpler *et al.*, 2012), for roads it is not that trivial because width of roads is almost never indicated in OSM shapefiles. Thus, many approaches pursue the challenging task of road pixels extraction supported by OSM. We can mention the approach of Yuan and Cheryadat (2013) where road shapefiles are projected into the aerial image. Then, extraction of homogeneous segments is performed. The distances from the upsampled vertices of OSM polygonal chains to the segment borders are computed. The clusters over distance values yield the road width. This approach is very elegant, however it does not tolerate roads with changing width, shadowy regions close to road borders, and inaccuracies in rotation of geo-referencing. More general is the approach of Mattyus *et al.* (2015), who apply aerial images as well. By utilization of the OSM data and various features extracted from the aerial image, an energy function is composed

and minimized to define the centerline of a road. Thus, OSM data is merely one of the components of the energy function that besides, reflect presence of vehicles, geometric consistencies, etc. Energy minimization reminds simulated annealing, which is known to be costly. In Lopes *et al.* (2017), OSM data is used to create special LCZ-based classes which are divided in urban, meaning different building types, and rural land cover, such as vegetation or soil. Additionally, OSM data were employed to eliminate classification inconsistencies. Those OSM data, which are represented by open polygonal chains, are attached with a buffer. Beside OSM shapefiles, WUDAPT data, Mills *et al.* (2015), are included for a more precise differentiation of the OSM data types. Here, sensor data are not really exploited, but the results are projected into a Landsat image. A combination of fuzzy pixel-based and object-based approaches for image classification is presented by Shackelford and Davis (2003a). In the first step, a fuzzy pixel-based classifier, analogous to maximum likelihood, based on spectral and spatial information for discrimination between urban land cover classes is applied. Pixels are evaluated with confidence values. Next, regions with high confidence values function as pre-classification parameters for supervised object-based classification. Even here, shadowy areas that cannot be assigned to any land cover class are denoted unclassified or added to shadow class. Consequently, a lot of important information is lost. In the contribution of Huang *et al.* (2015), the authors use OSM data to acquire training data for roads, but they also rely on differential morphological profiles (DMP) for buildings and shadows, normalized difference indexes for vegetation and water bodies, as well as HSV channels for soil (low Hue, high Saturation, and low Value). In this approach, areas obviously belonging to a class are easily detected using thresholds, and these same thresholds play an important part during upcoming classification. However, the difficult image regions, such as borders between classes (extremely important for some learners), or shadows, are ignored. This is a disadvantage of the algorithm and the reason why the authors (1) treat the shadow class separately, as a proper land cover class, (2) delete labels of pixels close to borders, and (3) conclude that active learning, that is, cautious selection of training examples, greatly improves the results. Our approach is different in the sense that we wish to perform the separation of classes in “easy cases” by means of few features only, and taking into account the difficult cases within an interactive step. Moreover, our features stem from image *and* elevation data while in classification, we go one step further performing non-local optimization. The research topic of Kaiser *et al.* (2017) is semantic segmentation of high resolution images, for each pixel a class label is assigned using supervised classification. Open image and map data are used as input for Convolutional Neural Networks (CNNs). Training is performed with OSM labels for building and roads as well as aerial images. This paper has two important conclusions: First, if the amount of data is extremely high and the learner is “data-hungry” (citation), the data does not have to be thoroughly prepared in order to achieve excellent results for the classes building and road. For example, initial road width values are parameters of road type only. In our work, we will show that after a minimum of manual interaction, it will be possible to obtain similarly good results from combined image and elevation data using only some 10% of the overall available labeled material and differentiating between classes tree and grass. Another general conclusion is that huge databases are very helpful for land cover classification by means of CNNs, which are employed for this task with rapidly increasing popularity and success (Kaiser *et al.*, 2017, Badrinarayanan *et al.*, 2015, Long *et al.*, 2015, Maggiori *et al.*, 2016).

2. <https://geo.gob.bo/geonetwork/srv/esp/catalog.search/-/metadata/dcb4efab-cf41-40aa-abde-a9f21b>

3. <https://www.geo-wiki.org/>

However, for two reasons, we decided to postpone the accuracy assessment of this classifier for our future work. First, since deep neural networks use multiple pooling layers, the results are computed at a lower resolution. Upsampling to the initial resolution is a crucial step in which the major scientific effort has to be invested. This step was not performed yet. Second, the selection of suitable feature sets does not make much sense for CNNs, but this is what we wish to investigate in this paper.

We complete the literature review on land cover classification by discussing those approaches, in which training data from open sources plays no or only an insignificant part and which, nevertheless, are quite motivating for our work. An example of an approach combining 2D and 3D data is presented by Zhang *et al.* (2015). The procedure is based on morphological top hat profiles applied to the brightness transformation of the orthophoto image, its negative, as well as to the Digital Surface Model (DSM). The authors differentiate between *top hat by erosion* and *top hat by reconstruction*, which are needed for a clean segmentation of off-terrain objects and accurate detection of slanted surfaces, respectively. Besides, the size of the structure element for the morphological operations is determined using training data, which probably only will work if the training data is object-wise complete. Guo *et al.* (2011) do not compute the DSM at all but apply their features and classifiers to the laser point clouds directly. This has the advantage that very useful features based on the absolute and relative echo numbers can be taken into account additionally to the usual ones, such as planarity and relative z-range. This last feature helps masking out ground points, and, in equal measure as DMPs used by Zhang *et al.* (2015), the explicit step of terrain modeling is not longer necessary. However, we already mentioned in the motivation that our goal is a geo-specific 3D model of urban terrain, so this step must be accomplished anyway. On contrary, Sun *et al.* (2017) have added both absolute and relative elevation as features. Another interesting aspect of their work was applying some of the Gray Level Co-occurrence Matrix (GLCM) features to the DSM and also the post-processing, for which CRFs were considered. The SVM (Support Vector Machine) ensemble approach of Huang and Zhang (2013) relies on three sets of features: The contrast (GLCM) feature, sampled over different filter sizes and directions, Urban Complexity Index (Yoo *et al.*, 2009), and DMPs. An SVM applied to each of these feature sets yields an output and confidence according to which pixels are classified. This process is supported by segmentation. A somehow surprising conclusion was that the best results were achieved for the strategy considering segmentation at the earliest stage. This, similarly to Wegner *et al.* (2015), motivated us to pursue a super-pixel-based approach. Critics may be raised that Huang and Zhang (2013), even though merely interested in segment-wise classification, had applied SVMs to every pixel feature vector, instead of concatenating features within the superpixels. Another conclusion of Huang and Zhang (2013) was that the accuracy of a certain feature set depends on the dataset. Therefore, more recent works (Montoya-Zegarra *et al.*, 2015; Huang *et al.*, 2017) presuppose concatenation of features and choice of classifiers robust to highly redundant features. In Maas *et al.* (2017), class labels from a map including classes absent in the new data are used for training. The big advantage is that efforts for acquisition of new training data are not indispensable anymore. However, robustness to noisy training data must be guaranteed in order to obtain an accurate transition matrix between the old and new classes. This is accomplished within an iterative workflow containing re-training and re-classification. Classification of satellite images into land cover classes is the topic of Fröhlich *et al.* (2013)

and is carried out fully automatically using segmentation and classification for each pixel without human interaction. To provide separation between land cover classes, contextual information is employed. Beside the satellite image, elevation data in form of a normalized DSM could be derived and used as further channel. Here, Iterative Context Forest is applied for classification.

Background

This section is divided into two parts: The first part describes the input data and its main pre-processing steps, and the second part provides an overview about segmentation methods considered for land cover classification and about how to operate with these segments.

Data Preparation

The data we dispose of is the co-registered airborne optical and elevation data. The former is given by a true orthophoto while the digital surface model (DSM), which can be sampled from the 3D point cloud, gives the latter. The point cloud can be captured by an active airborne sensor (ALS) or by 3D stereo reconstruction. Small holes (uncovered areas) in the digital surface model (DSM) can be closed by e.g., harmonic inpainting with a boundary condition (Neumann). In hilly terrain, it is more reasonable to work with relative rather than with absolute elevations. Therefore, computation of the Digital Terrain Model (DTM) by means of e.g., Bulatov *et al.*, 2014. Sec. 2.1 constitutes the next important module of data preprocessing. The difference between DSM and DTM yields the relative elevations, usually denoted as Normalized Digital Surface Model (NDSM).

In order to define suitable target classes, we observed that Kaiser *et al.* (2017) were merely interested in buildings and roads, defining the remaining land cover classes to be background. Also for us, the successively applied algorithms of building reconstruction, e.g., (Bulatov *et al.*, 2014; Sohn *et al.*, 2012) and road centerline extraction (Mena, 2006; Bulatov *et al.*, 2016a) are the main motivation to achieve good classification results. Additionally, we are able to differentiate between high (tree) and low (grass) vegetation; first, because elevation channel is available, second, because GIS-data occasionally include large forest regions and positions of tree crowns (denoted as natural points), and, finally, because algorithms of individual tree detection and representation (Straub, 2003; Bulatov *et al.*, 2016b) are also important for our urban terrain models. Further classes could be bare soil and water bodies. Here, Huang *et al.* (2015) had already demonstrated successfully that considering further indexes, such as those stemming from the HSV color space, training data can be acquired for these classes, too. However, in urban terrain these and further classes, such as burnt land, waste disposal sites, earth hills and trenches, etc., appear quite seldom while we want our data to be balanceable (that is, classes appear in coarsely equal measure over the urban terrain) and easily separable by as few measures as possible (for instance, relative elevation, vegetation index, planarity). Because of these reasons, we will merge grass, which is a class less interesting for us, with the clutter (bare soil, croplands, tennis courts, etc.). Analogously, pixels occupied by vehicles, road signs, etc. will be counted as impervious surface. Here, we can argue that the holes caused by detection of these smaller objects result in quite noisy courses of road centerlines, so that most of the road centerline extraction algorithms pre-supposed hole closing as a preprocessing module (Mena, 2006). Finally, we emphasize that the four chosen classes, building (roofs), road (more accurately, impervious surface), tree and grass, remain

widely constant within moderate periods of time, differently to vehicles and croplands.

Segmentation and Operations with Segments

For land cover classification in high resolution data, it is often worth to think about whether classification should take place pixelwise or segmentwise. Since many experiments with different parameter settings must be performed, we opt in this work for an approach in which labels are assigned to superpixels. Under field conditions, the main motivation to do so is that the result of a good superpixel algorithm widely respects borders of land cover classes. In other words, few superpixels contain non-negligible percentages of pixels belonging to several classes, and thus, changes of resolution are barely noticeable. At the same time, superpixel-wise performance accelerates immensely the classification step and often improves the results since pixels' larger neighborhoods are taken into account in a natural way during computation of features. Grouping of features into segments costs additional time, however, we will see how our further considerations allow to keep it relatively low.

An example of such a reliable tool for superpixel computation is the so-called *compact superpixels* algorithm proposed by Veksler *et al.* (2010). This segmentation is supposed to respect, on the one hand, the local changes of image gradient while on the other hand, too complicated border courses are avoided. These two constraints are integrated into a non-local energy minimization framework over the neighboring superpixels. Usually, a good local minimum of the energy function is quickly found by an iterative application of the alpha expansion algorithm (Boykov *et al.*, 2001). After the energy minimization, the result is slightly postprocessed in order to delete too small superpixels and to achieve a consistent labeling (Bulatov and Warmke, 2017).

The extension of a superpixel is limited by the initial grid size, which is a trade-off between the accuracy and the computation time. Typically, it should be slightly smaller than the tiniest object of interest in the dataset, such as the smallest building (part), narrowest pathway, etc. The limited size is a problem if we are interested to extract rather large homogeneous image regions for our training data acquisition module. This is why additionally to the superpixel decomposition, we applied the segmentation algorithm due to Wassenberg *et al.* (2009). This method is quite fast because the graph considered by Veksler *et al.* (2010) is simplified by the minimum spanning tree (MST). Besides, a heuristic similar to Canny's solution for edge extraction in images has been proposed. The larger the region s , the higher is the confidence that it corresponds to a true object. Hence, the more generous is the threshold according to which additional pixels are added to s . The weights of edges are given by the L_2 distance between color values of the corresponding pixels. Thus, an image with an arbitrary number of channels can be theoretically considered. However, in order to keep the input unchanged, we stick to the three-channel image, only the blue channel is replaced by the NDSM rescaled to the same order of magnitude as the colors of the orthophoto. This was done to prevent pixels with different elevations to belong to the same segment. The main reason to use the empirical value for scaling is that the accuracy of 3D reconstruction, which should ideally be reflected in this scale factor, is not always available, unfortunately. Since only large regions are considered as training examples, filtering of segments with respect to their area is successively performed.

The authors are aware that two different tools for image decomposition could be confusing for the reader. We will refer to *segments* at the stage of training data acquisition and to *superpixels* during classification. Note that every image pixel belongs to *exactly one* superpixel and to *at most one* segment,

since small segments are deleted. At the same time, the size of a superpixel is bounded while the segmentation algorithm tends to capture whole regions.

In the last paragraph of this section, we refer to operations with superpixels and segments. Starting at a pixel feature, for example, elevation $h(x)$, we want to assign features \bar{h} to a segment (analogously, superpixel) s as well. The evident way to do so is to concatenate the feature over all $x \in s$ by means of one or several statistical measures. In other words, we define

$$\bar{h}(s) = \bar{h}_\phi(s) = \phi_{x \in s}(h(x)), \quad (1)$$

where ϕ can be a scalar or vector-valued statistics measure. Examples are simple functions (mean, median, standard deviation, range), histograms with a fixed size of bins (Dalal and Triggs, 2005), or even multi-dimensional (dictionaries of) descriptors (Fulkerson *et al.*, 2009; Montoya-Zegarra *et al.*, 2015). To compute these measures efficiently, we propose the following approach: By keeping in memory the sorting strategy and the breaking points, the image pixels are stored according to segment or superpixel number as illustrated in Figure 1. Breaking points are thereby the cumulative sums of cardinalities. If a new set of features (as will be explained later) is available in form of a one- or multi-dimensional image, its pixels are accessed in the memorized order. Now, ϕ is computed column-wise and its computation is highly accelerated by parallel processing or using software optimized for matrix operations, such as MATLAB.

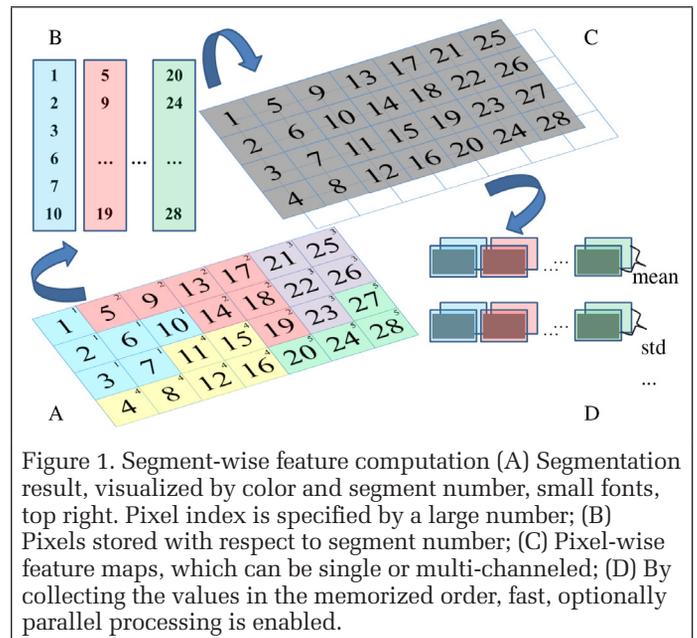


Figure 1. Segment-wise feature computation (A) Segmentation result, visualized by color and segment number, small fonts, top right. Pixel index is specified by a large number; (B) Pixels stored with respect to segment number; (C) Pixel-wise feature maps, which can be single or multi-channelled; (D) By collecting the values in the memorized order, fast, optionally parallel processing is enabled.

Classification

Starting from the combined optical and elevation data, we wish to derive those properties which enable a clear differentiation between the land cover classes in question. The most obvious idea is that trees have high values of both NDSM and NDVI, impervious surface (like roads) is characterized by low values of NDSM and NDVI, grass has high values of NDVI and low values of NDSM while for buildings, it is vice versa. However, an accurate choice of thresholds is difficult because according to Bruzzone and Carlin (2006), as well as Huang *et al.* (2017), a higher resolution contributes to an increasing intra-class variation and decreasing interclass variation. Additionally, in real datasets, many counterexamples exist which require additional model assumptions. For instance, bridges

and highway crossings have non-negligible elevations over the ground; building roofs are sometimes covered with grass. Therefore, we utilize these apparently most discriminative features for training data extraction only, as explained in the first subsection of this section, while for land cover classification itself, a more sophisticated set of features is developed. Next, we provide motivation and technical details for the set of features, both on pixel and superpixel level. Afterwards, we will provide introduction to both learning algorithms. Finally, we will describe the superpixel-wise non-local optimization module. Minimization of an energy function consisting of a data term from the classifier and a smoothness prior is supposed to enforce neighboring instances to belong to the same class. We illustrate the input data and the processing chain in Figure 2, while for the core procedure of training data extraction, Algorithm 1 is provided.

Training Data

In order to obtain sufficient examples of training data, we apply OSM data in form of shapefiles. The shapefiles are vector data, and thus must be rasterized. In case of buildings, it is a straightforward task to check whether a pixel lies inside of a closed polygonal chain. In case of streets, the situation is less clear, because a road segment is basically given by a polygonal chain and the road width is not part of the shapefile attributes. One helpful attribute is road type (residential, secondary, cycleway, forestry, steps, etc.). However, no general conclusions can be drawn about the road width. Consequently, we only used this attribute in order to suppress too small and insignificant roads and, in order to rasterize the road data, we rely on the segments (Wassenberg *et al.*, 2009), taking into consideration those which have an intersection with polygonal chains; to find such segments, we rasterized the latter ones with one pixel width. This approach may remind that of Yuan and Cheryadat (2013), but is carried out for a different purpose. Things are becoming even more tricky in case of single trees and grass areas. While they are often not part of the OSM data at all, only sometimes in urban centers trees are stored as single point positions in UTM coordinates and one could take a segment around these points. However, usually because of changes in elevation and presence of high-level textures (branches, leaves, shadows), only a few small segments are yielded. Furthermore, we need examples for groups of trees within forest regions, which are, in general, not contained in OSM data. Therefore, and also because it requires a closer look and botanical skills to make a difference between tree and grass, we decided to threshold the orthophoto with respect to the relative elevation and NDVI, to perform connected component analysis, and finally to mark several larger areas of high and low vegetation. However, this should be done after segments for buildings and roads have been set, since otherwise we would overwrite the information provided by OSM.

The OSM data are not error-free. We found out that there are three important error types: (1) Buildings are often generalized; inner courts are considered to be part of the building while they mostly belong to asphalt surface or grass; (2) Roads are sometimes occluded by buildings, tree crowns, or are overgrown with grass; and (3) Trees turned out to be subject of very frequent changes, especially decorative plants near squares in urban centers. Besides, the positions indicated in OSM shapefiles are not always accurate and especially in winter, they correspond to an asphalt or grass point quite often.

To pursue our plan to correct these errors by means of the actual sensor data, it is helpful to look at the problem from the statistical point of view. We formulate model assumptions on classes and evaluate them using sensor data. To a greater extent, the four classes can be distinguished by means of the two most substantial variable, NDSM and NDVI only, leaving

aside additional important features, like planarity (Gross and Thönnessen, 2006) or saturation, defined as $\max(R, G, B) - \min(R, G, B)$, where R, G, B are red, green, and blue channels of the orthophoto. This means that the so-called second-type errors (or errors of the second kind), mentioned in the previous paragraph, can thus be detected and deleted from the pool of examples. For instance, masking out all segments with the average elevation exceeding 1 m and those having the average NDVI value exceeding a conservative threshold would indeed result in segments belonging to the road class only. For the threshold, one can roughly estimate how much of the image belongs to the vegetation class and select the quantile value according to this percentage. Analogous argumentations can be provided for the other three classes as well. Even with these two thresholds, the remaining errors of the second kind are already seldom; for example, if a yellow autumn tree occludes a building roof. It is declared as a building by OSM data and the hypothesis would be spuriously accepted by considering NDSM and NDVI. To prevent this from happening, multiple thresholds imposed on NDSM, NDVI, planarity, and saturation, are replaced by a classifier, such as Random Forest. Hereby, our trick to compensate for the different segment sizes is simply to utilize more segment-based statistical measures from a previous section: Taking mean, median, variance, minimum and maximum values over each segment, we obtain a feature set consisting of several dozens of features. Since we want to exclude the errors of the second kind, we consider merely segments with a high probability to belong to either class.

However, doing so increases the risk of errors of the first kind (rejection errors). A non-accepted hypothesis could be a bridge, a road segment in shadowy regions, a dead or leafless tree, etc. Deleting the corresponding segments could have the dangerous consequence that only easy examples would be found. This means that during the upcoming classification step, assignment of difficult examples to a class would be more or less random because the classifier would ignore every other feature except those used for detection of errors of the second kind. We display the temporarily rejected segments, denoted by B , R , G , and T in Algorithm 1, in our user interface. Out of those representing errors of the first kind, a few can be assigned to the correct class by means of a few mouse clicks. This step is interactive; however, we have experienced that during training data acquisition, minimum user interaction is indispensable, at the latest, for quality control. Therefore at this stage, we considered it convenient and recommendable to provide a few additional examples instead of thinking out more complicated rules. The advantages are: Savings of resources, predictable expenditure of time, as well as the fact that this task can be carried out by a non-professional.

The final step is postprocessing. Remember that we deleted most of the small segments, such that instead of chimneys, street marks, traffic lights, etc., the unlabeled holes remain in the segmentation result. These holes and other high-frequency oscillations near the segment borders are closed by means of morphological operations.

Pixel-Wise Features

The most obvious feature of a pixel set is the pixel information itself. The initial set of features, to which we will refer to from now on as *unfiltered*, is comprised of the raw channels of the orthophoto as well as of the relative elevation. More sophisticated combinations of channels are possible. We add to this set the NDVI measure involving the near infrared and red channel, if the former is available. Otherwise or additionally, also the pseudo NDVI measure, in which near infrared is replaced by green, is a good measure to distinguish the regions occupied by vegetation. As the last measure for unfiltered features, we take the saturation channel of the

HSV representation of the orthophoto. Here, we already see a certain redundancy of the chosen features since, for example, the red channel can be extracted by means of green and the pseudo NDVI measure. However, doing so we hope to obtain a certain stability in the set of features. The negative effects of the redundancy should be compensated first by the feature selection module and second by a classifier which has proven to be rather stable with respect to redundant features, such as RF (Genuer *et al.*, 2010).

Additionally to the unfiltered pixel features, we wish to consider the local neighborhood of the pixels in order to identify the local structures and patterns, such as edges, scratches,

Algorithm 1. Overview of the algorithm for training data extraction. *wrt* = with respect to and \sim denotes complement. See text for further details.

-
- Step 1. Rasterize OSM buildings to obtain binary image B . Filtering wrt NDSM, NDVI and planarity derived from sensor data yields confirmed B_+ and non-confirmed B_- buildings.
- Step 2. Rasterize OSM-based road centerlines with segments of Wassenberg *et al.* (2009) to obtain binary image R . Fusion of larger segments with \tilde{B} and with thresholded NDSM and NDVI images yields confirmed R_+ and non-confirmed R_- road segments.
- Step 3. **if** exist OSM data for tree and grass areas
 // forests and meadows
 Fusion with sensor data, as well as with \tilde{B} and \tilde{R} yields confirmed G_+ , T_+ and non-confirmed G_- , T_- grass and tree areas.
 otherwise
 Filter orthophoto by NDVI.
 Thresholding wrt NDSM and planarity yields G_+ , T_+ , G_- , and T_- .
 end if
- Step 4. Optionally: interactive selection of training data from B_- , R_- , G_- , T_- , and remaining unlabeled pixels.
- Step 5. Morphological post-processing and connected component analysis. Output: B_+ , R_+ , G_+ , and T_+ .
-

blobs, etc. Usually, this is accomplished by applying filters which are concatenated within filter banks (Cula and Dana, 2004; Varma and Zisserman, 2005; Winn *et al.*, 2005). We opted for the filter bank proposed by Varma and Zisserman (2005). Here, the edge and bar filters are applied in different rotations and then the maximum filter response among the rotations is taken in order to make the filter bank rotationally invariant, which is important for our task. Similarly to Warnke (2017), Wegner *et al.* (2015), we applied the filters contained in this filter bank to the infrared channel, to the first channel of the Opposite Gaussian Color Space transformation (Geusebroek *et al.*, 2001), and to the elevation map.

The aforementioned filters are not always able to capture objects of arbitrary size. The morphological operations of path opening and closing with different filter sizes allow to detect respectively to suppress fair and dark blobs of different extensions in the image. Because they enable view-point and contrast-invariant analysis of textures, they are tried-and-true techniques in Remote Sensing. Therefore, we applied differential morphological profiles (Pesaresi and Benediktsson, 2001) of sizes 1, 3, 5, 10, and 15 meters to the first channel of the orthophoto and the DSM. Besides, we computed from the DSM the planarity measure, using the three eigenvalues of the so-called structure tensor in the square neighborhood of a pixel. The square size in pixels is 2 if the resolution of the data exceeds 0.5 and 1 otherwise. From the first channel of the orthophoto, we computed entropy filters using two different kernels.

Finally, we present features based on stripes, which is an example of higher-level variables. The motivation to use these very context-based features could be derived by one of the conclusions of Huang and Zhang (2013): Commonly used feature sets, such as DMPs, GLCM, etc., exhibit varying performance for different data, because some react on blobs, others on color combinations, etc. Overall, it is very difficult to make a prediction about the performance of a certain feature set. However, one could consciously search for pairs of parallel lines in the image, and use this information about their presence, length, distance, etc., as features for superpixels between the parallel lines. Thus, if there are many challenging rectangles in the data, like shadowy street canyons, complicated piece-wise rectangular roofs, dirty or grassy roads, etc., it is appropriate to apply stripes.

To our knowledge, stripes have been first used by Soergel *et al.* (2006) for detection of airports and bridges. Then, Schilling *et al.* (2015 and 2018) applied stripes and relevant features for vehicle detection while Bulatov and Warnke (2017) used them for detection of roads. In the latter paper, the authors carried out a classical feature selection and it turned out that in the Vaihingen dataset (see the next Section below), stripes were among the most important features. They helped to extract information which goes beyond the superpixels' borders, and thus to perform reliable road detection in those dark shadowy regions where most of the features previously discussed are not distinctive enough. All this is consistent with the considerations we made in the previous paragraph and served for us as a motivation to employ stripes for multi-class problems for detection of man-made structures.

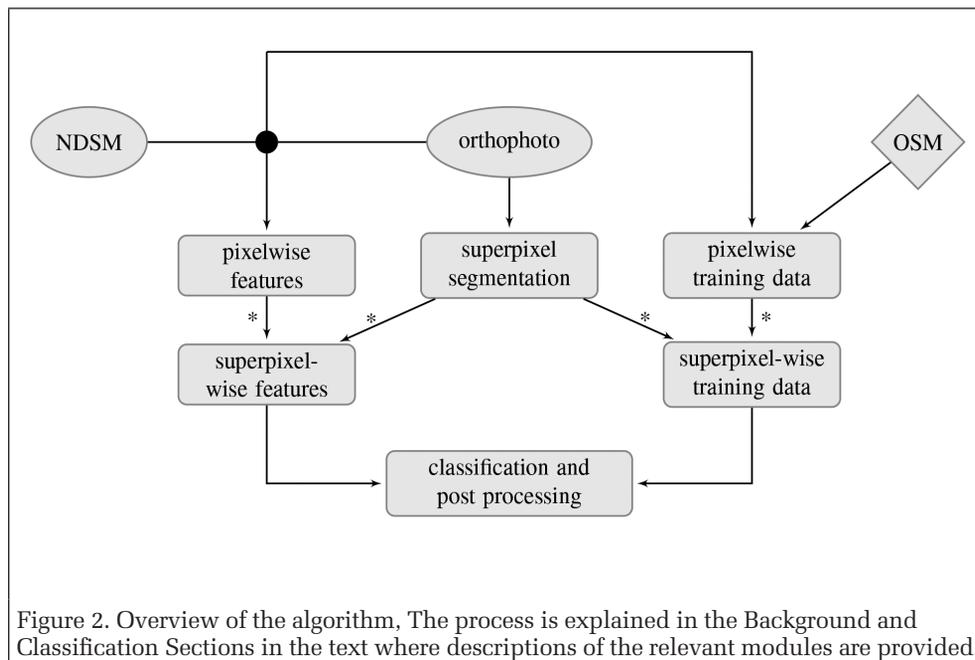


Figure 2. Overview of the algorithm, The process is explained in the Background and Classification Sections in the text where descriptions of the relevant modules are provided.

This is why in this paper, we will separately treat street-like and building-like stripes. Straight line segments are computed by means of the Burns algorithm (Burns *et al.*, 1986) in ortho-photo and DSM. They can be prolonged in the case they have close endpoints and similar orientation. Stripes are quadrilaterals whose vertices are endpoints of pairs of nearly parallel line segments with a sufficient overlap along the dominant direction. A road-like stripe has a low average NDVI value and negatively outstanding elevation:

$$\bar{h}(S) < \min(\bar{h}(S_-), \bar{h}(S_+)), \quad (2)$$

where S_- , S_+ are rectangles formed on both sides of the parallel lines of this stripe and \bar{h} is from Equation 1 with mean value function for ϕ . Conversely, a building-like stripe has a low NDVI value and a positively outstanding average elevation, resulting in a trivial modification of Equation 2. Note that stripes are always convex polygons that can be quickly rendered into the image after which the pixels inside of a stripe are stored. Hence, computation of e.g., $\bar{h}(S)$ takes places in the similarly efficient way as previously described and illustrated in Figure 1, only that a pixel can belong to more than one stripe.

Segment-Wise Features

The step from pixel-wise to superpixel-wise features is conceived as previously explained: The feature value of a segment results from concatenation of that of its pixels by one or several statistical measures. For the features not based on stripes, we decided to restrict ourselves to the mean value and variance. This choice, already preferred by Wegner *et al.* (2015), has achieved a good performance. The number of stripes covering a region is the maximum function of the pixel measure *sum of stripes containing a pixel*, where the sum can, additionally be weighted by how building- and street-like the stripes are. To assess this, their extensions are taken into account, too. Finally, the percentage of pixels occupied by stripes is the mean value of the logical measure pixel covered by stripes.

Learning Algorithms

Including a high amount of features, as described before, may lead to redundancies. Moreover, some are not even relevant for a particular dataset or class. A feature evaluation will help to enable a stable classification and to choose the right feature set depending on the used classifier. Fortunately, our decision to use Random Forest classifier (Breiman, 2001) proves to be beneficial since it is an ensemble learner which works stably despite a certain amount of redundant or non-relevant features. Additionally, we considered the Import Vector Machine which is a hyperplane-based classifier, similarly to the far better explored Support Vector Machine. Contrary to SVM, whose output needs an additional transformation to reflect probabilities, IVM has a probabilistic output we will need in the following section. Because of not-trivial functioning of this classifier, we considered a short introduction about IVM to be of interest. Also, an analogous overview about RF will be provided as well. Our motivation here is that RF has a relatively simple underlying algorithm, however, its mathematical properties, e.g., with respect to convergence, are more or less unknown (Biau, 2012). Throughout this section, we use the following notation: The training set is denoted by $(x_n, y_n) \in \mathcal{T}$. It comprises N feature vectors $\{x_1, \dots, x_n, \dots, x_N\}$ and corresponding class labels $y_n \in \{1, \dots, c, \dots, C\}$. The feature vectors are collected in a feature matrix $X = [x_1, \dots, x_N]$.

Random Forest: The algorithm basically consists of two steps, which are repeated for several iterations I_{max} . In every iteration, a fixed number n of training examples from X with the corresponding labels are sampled, whereby replacements

are permitted. For such a subsample, a classification tree f_j is built by successively finding optimal splits in small and random subsets of features, and then increasing tree size using the bootstrap aggregation techniques. The key idea of these techniques is that using repeated samples, and thus correlating the trees, conclusions can be drawn on how separable or discriminative any single feature is.

To classify an unlabeled point (superpixel s), probabilities of predictions by means of the trees are required. In our case,

$$p_c(s) = p(y(s) = c) = \#(f_j(s) = c) / I_{max}, \quad (3)$$

where $\#$ denotes cardinality. One can imagine this classifier as a union and intersection of many hyper-rectangles. It is intuitively clear – and for several simplified versions of Random Forest it can be proved analytically – that it is quite robust to many redundant and even irrelevant features (Biau, 2012; Genuer *et al.*, 2010).

Import Vector Machine: IVM is a discriminative and probabilistic classifier, realizing sparse kernel logistic regression. It has been introduced by Zhu and Hastie (2005) while Roscher *et al.* (2012) have shown that IVM provides reliable posterior probabilities. To account for complex decision boundaries between classes, the original features X are implicitly mapped from the input space to a higher-dimensional kernel space, resulting in a kernel matrix $K = [k_{nm}]$, $n, m \in 1, \dots, N$ obtained by applying a kernel function $k_{nm} = \kappa(x_n, x_m)$. We utilize IVM with a Gaussian radial basis function kernel parameterized by kernel width σ , which is commonly used for remote sensing purposes. Since using the whole kernel matrix is too computationally expensive, similar to SVM, an IVM chooses a subset v of feature vectors out of the training set with $V = |V|$ samples $X_v = [x_{v,m}]$, $m = 1, \dots, V$, obtaining a sparse solution of the kernel logistic regression. These feature vectors are called import vectors. The objective function is given by

$$\mathcal{Q} = -\frac{1}{N} \sum_n t_n^T \log p_n + \frac{\lambda}{2} \sum_c \alpha_c^T K_R \alpha_c \quad (4)$$

with probabilities $p_n = [p_{n1}, \dots, p_{nc}, \dots, p_{nC}]$ obtained by

$$p_{nc} = p(y_n = c | k_n; A) = \frac{\exp(k_{v,n} \alpha_c)}{\sum_l \exp(k_{v,n} \alpha_l)}. \quad (5)$$

The parameters $A = [\alpha_1, \dots, \alpha_C]$ parameterize the multi-class decision boundaries, and can be determined using, for example, the Newton-Raphson procedure. We add an L_2 -norm regularization term with parameter λ to prevent overfitting. The V -dimensional kernel vector is given by $k_{v,n} = [\kappa(x_n, x_{v,1}), \dots, \kappa(x_n, x_{v,V})]$ and the $(V \times V)$ -dimensional regularization matrix is defined by $K_R = [\kappa(x_{v,1}, x_{v,m})]$, $\{l, m\} = 1, \dots, V$. The binary target vector $t \in \{0, 1\}$ of length C uses the 1-of- C coding scheme such that all components but t_{nc} are 0 if the feature k_n is from class $y_n = c$. For a more encompassing description of IVM, we refer to Zhu and Hastie (2005) and Roscher *et al.* (2012).

Non-Local Optimization

The advantage of both selected classifiers is that the scores assigned in Equation 3 and in Equation 5 for RF and IVM, respectively, for a superpixels' label can be directly associated with the likelihoods. However, this local result may appear noisy because neighborhoods are not considered. Thus, priors are taken into consideration. Generally, neighboring instances (pixels or superpixels), which do not belong to the same class, are penalized. This strategy is pursued in this work as well. We interpret labels of segments as random variables in a Markov Random Field and consider the standard energy function

$$E(c) = \sum_s -p_c(s) + \alpha \sum_{s,s' \in \mathcal{N}} U(c_s \neq c_{s'}), \quad (6)$$

where $p(s)$ is the data term incorporating the probability of the classification result (for the sake of simplicity, we omit here the frequently used negative log-likelihood), $\alpha > 0$ is the smoothness parameter, U is the indicator function, and \mathcal{N} is the neighborhood system, which in the case of superpixels, is easily established from the non-zero gradient map of the superpixel image. From Equation 6, one can notice that we treat all classes symmetrically, penalizing unequal labels for a pair of neighboring pixels by a constant, independently on these labels. This may result in less accurate assignments, e.g., of the classes *building* and *road*, than in the approach of Montoya-Zegarra *et al.* (2015), where minimum cost paths along road pixels or, respectively, alpha-shapes for buildings, were taken into account. However, we wish to keep our algorithm stable and without necessity to fine-tune numerous parameters. Additionally, we do not want to bias the evaluation results too much and therefore decided to maintain as few constants as possible.

To minimize the energy function, any of the methods mentioned in (Szeliski *et al.*, 2006) can be applied and the results are rather similar because of the simplicity of the prior. We chose the graph-based method of alpha expansions due to Boykov *et al.* (2001), for which we used the freely available implementation of Delong *et al.* (2012). Due to a small number of labels, it converges after a few iterations. We noticed that the performance of other optimization algorithms is similar to the chosen one. It should also be mentioned that for pixel-based approaches, the semi-global method of Hirschmüller (2008) is applicable, too, as was demonstrated, e.g., by Schindler (2012). Because this method exploits the grid-like structure of images, it is fast and, therefore, widely used in remote sensing.

Experiments

In this section, we will provide a detailed description of our results. We wish to analyze the impact of different feature sets, depending on whether the data stems from a rural or urban area, whether the elevation information stems from an airborne laser scan or has been retrieved using stereo matching method, and whether the data for training and validation was selected as previously described or annotated manually. Therefore, we decided to cover a broad spectrum of datasets, which we present at the beginning of this section. The second subsection is dedicated to the evaluation strategy while in the final subsection, classification accuracy will be reported.

Datasets

The first dataset is the ISPRS benchmark dataset *Vaihingen*, provided by the German Society for Photogrammetry, Remote Sensing and Geoinformation (DGPF). *Vaihingen* is a small town in hilly area of Southern Germany. It has some industry and other quite large buildings, residential areas with complex roof structures, vineyards, as well as roads in non-negligible altitudes over the ground. Originally images of GSD 0.1 were used to obtain 3D information (Haala and Rothermel, 2012). The 3D data (with a reported accuracy of 0.12 to 0.16 m) was then sampled into a DSM while the three channels (red, green, and near infrared) of the image data were transformed into a true orthophoto. The data was subdivided into 37 patches of different size and complexity. To 16 of them, a ground truth was provided by the Siradel corporation as well as Spreckels *et al.* (2010).

The second dataset represents an inner part of the city of *Munich*. The procedure for obtaining the DSM and the true orthophoto was the same as in the dataset *Vaihingen*. It was provided by the Institute of Photogrammetry of Stuttgart (IfP). The image had four channels and we downsampled the data by factor 2 in order to accelerate the evaluation. Contrary to the first dataset, *Munich* is flat. Most buildings are rather high and have a multi-facetted geometry and texture properties of

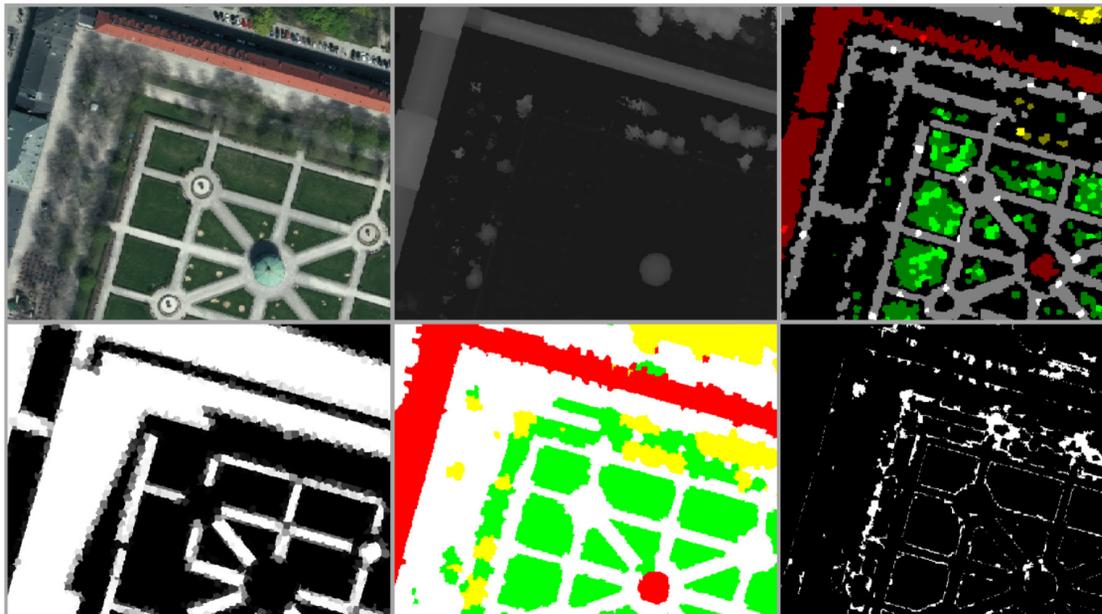


Figure 3. Input data and qualitative results for the dataset *Munich*, park area. Top row: Input data; corresponding fragments of the orthophoto (left), NDSM (middle), and labeled data for classes building (dark-red), street (grey), grass (dark-green), and tree (dark-yellow). Superpixels used for learning are highlighted (red, white, green, and yellow). Bottom row, left to right: Road-like stripeness features for the orthophoto (left), superpixel-wise classification result of Random Forest with all features and non-local optimization (middle), and pixel-wise evaluation result for this configuration. Correctly and incorrectly assigned pixels are denoted by black and white color, respectively.

roofs. Another consequence of high buildings is abundance of shadowy areas, such as street canyons, in which we expect our stripeness measures to be successful, because main streets exhibit stripe-like patterns, see, for instance, Figure 3. Finally, both leafless and leafy trees are present in the dataset, reflecting in difficulties caused both by under-segmentation and during training data acquisition.

The last dataset is from a quite rural area in the outskirts of the Northern German city of *Münster*. However, in the Eastern part of the dataset, there is a residential area. The aerial image at the original resolution of 0.1 m was downsampled to fit the airborne laser scan which was resampled to the resolution of 0.5 m. Note that rescaling of images was merely performed for the sake of convenience and is, in general, not necessary: For a superpixel-based approach, features can be rendered at different resolutions. Also here, there were four channels. The main challenge of this dataset is the low resolution. Streets and paths in the rural part of the dataset only have a width of several pixels. The texture properties of these paths strongly remind patterns in croplands, which are also present, however, according to our convention of above, are merged with the low vegetation class.

We provide an overview about all datasets in Table 1. From this table, one can estimate the total numbers of pixels and superpixels as well as the percentage of labeled ones. In the next section, we will explain how superpixels are labeled and evaluated.

Evaluation Strategy

In this section, we discuss the evaluation strategy, which is made up by the ground truth (training and validation data), evaluation metrics, as well as the actual goals of the experiments. First, we need to define our ground truth. Pixel-wise labeled maps are only available for *Vaihingen* patches for the classes building, road (impervious surface), grass, trees, car, and clutter. The class car was merged with the road class while pixels belonging to clutter remain unlabeled. Since the ground-truth result is given pixel-wise and we wish to assign classes to superpixels, we define as labeled only those superpixels, which entirely lie in one class (no unlabeled pixels). Because no ground truth were available in *Munich* and *Münster* data, we labeled large areas following the previous procedure of, performed training on a small portion of the labeled data, and used the remaining data for validation of our method. The training data was additionally balanced, that is, for each class we took an approximately equal number of examples, 30% of the least frequently occurring class. It turned out that only 5 to 15% of all labeled superpixels were employed for training and the rest for validation. In Figure 3 and in the images provided at the end of the Evaluation Section, the labeled data and that used for training (highlighted) are shown. The top right image of Figure 3 gives an impression of how imbalanced the labeled data is: For example, there are so many buildings in Munich in comparison to the tree class that only two superpixels of the whole huge building with the red roof in Figure 3 were selected for training. As one can see, we considered for learning superpixels randomly distributed in the images. In other publications, training and test data were separated, however, for performing classification in time-critical, realistic scenarios, one would rather tend to obtain examples from different image areas. Since

the results did not vary that much and the methods of active learning will be left out of consideration for this work, we accepted a certain randomness of the output.

Coming to the evaluation metrics, classification accuracy is conventionally measured by means of confusion matrices. For each parameter setting, we thus track the percentage of instances lying off the main diagonal of the confusion matrix. The validation was also performed for the training data, which is useful to measure over fitting. Since eventually, a classification result is required for pixels, we employed the winner-takes-all strategy for validation. This means that all pixels within the segment were labeled according to the most frequent class and thus, we accepted pixel-wise errors inevitably emerging due to under-segmentations. The error assessment on the pixel level was carried out for the local and the best – with respect to superpixel-wise classification – non-local result by means of confusion matrices on the pixel level. In Tables 2, 3 and 4, percentages of incorrectly assigned segments, training segments, and pixels are noted in columns as s., tr. s., and pix., respectively.

The main goal of this paper is to compare the accuracies measured by different classifiers, (groups of) features and non-local optimization. A model is learned for every feature set for Random Forest and Import Vector Machine. For the former classifier, the total number of trees was 40 is reasonable because from here number on, there was no significant decrease of out-of-bag errors. Moreover, the default number of variables per tree was \sqrt{N} , rounded, where N is the total number of variables. For the latter classifier, it is recommendable to normalize the data to zero mean and unity standard deviation. Moreover, an extensive search of the two key parameters, namely kernel width σ and regularization constant λ in Equation 4, is performed within an exponentially-equally spaced grid to guarantee fairness of the comparison. The configuration of σ and λ yielding the best results with respect to cross-validation is evidently chosen.

While for performance testing of the non-local optimization, one should measure the performance for varying values of smoothness parameter α in Equation 6 for different configurations, with respect to the selection of features, we originally wanted to perform classical feature selection (Bulatov and Warnke, 2017). However, for two reasons, this strategy turned out to be non-feasible. First, the maximal number of features (>100) is rather high. For Random Forest, it has already been mentioned that the influence of redundant features is not substantial; moreover, the features are even ranked according to importance within the algorithm. For Import Vector Machine, extensive search for the best values of σ and λ should be performed for every iteration of a feature selection algorithm, that is, a quadratic number of times. This would make computation time intractable even on the superpixel level. Second, most image operations are typical for every feature group and picking single features from any of the groups does not necessarily bring a desired reduction of computation time. Instead, one could wonder whether the pre-defined feature types, such as morphological profiles, filter banks, stripes, variances of segments' features etc., are indeed necessary. The choices of filters we consider are:

Table 1. Overview of the datasets. SP means superpixels while SP size corresponds to initial grid size in pixels.

Dataset	resolution	image size	origin 3D data	num. labeled SP	SP size
<i>Vaihingen</i>	0.1	$\approx (2 \times 2)$ MP $\times 3 \times 16$	stereo	363833	10
<i>Munich</i>	0.2	$\approx (4.7 \times 6.0)$ MP $\times 4$	stereo	80678	10
<i>Münster</i>	0.5	(2.4×4.1) MP $\times 4$	laser scan	133343	5

Table 2. Relative values (in %) of wrong assignments for the dataset *Vaihingen* for the local and the non-local method with the superpixel-wise best choice of smoothness parameter. For each setting, three results are reported: Superpixel-wise wrong assignments for all labeled data (s.), wrong superpixel-wise assignments for the data used for learning (tr. s.) and the pixel-wise wrong assignments (pix.).

feat. set	num. of feat.	RF						IVM					
		loc.			non loc.			loc.			non loc.		
		s.	tr. s.	pix.	s.	tr. s.	pix.	s.	tr. s.	pix.	s.	tr. s.	pix.
1	8	8.7	0.1	16	7.3	3.5	14.6	11.1	11.2	17.8	10.5	10.6	17.4
2	16	8.2	0	15.4	6.8	3.4	14	12.6	12.8	19.2	11.8	11.9	18.5
3	16	7.7	0	15	6.6	2.7	13.9	24.2	24.2	29.8	23.8	23.7	29.5
4	38	7.7	0	14.9	6.5	2.8	13.7	9.6	9.6	16.3	8.6	8.6	15.6
5	58	6.7	0	13.9	5.6	2	12.9	17	17.1	23.1	14.7	14.8	21.6
6	72	10.8	0	17.6	8.5	3.1	15.7	25.7	26.1	31.8	24	24.2	30.4
7	104	7	0	14.2	5.9	1.9	13.2	9	8.9	15.8	8.3	8.2	15.2
8	76	6.9	0	14	5.8	1.8	13	8.7	8.6	15.6	8	7.9	15.1
9	76	7.6	0	14.8	6.4	2.6	13.7	9.7	9.3	16.4	8.3	8.3	15.4
10	96	11.2	0	18.8	9.1	2.6	17.1	14.9	14.6	21.8	12.6	12.6	20
11	108	6.9	0	14.2	5.9	1.9	13.2	9.4	9.4	16.3	8.6	8.8	15.8

Table 3. Relative values (in %) of wrong assignments for the dataset *Munich*. For each setting, same results as in Table 2 are reported.

feat. set	num. of feat.	RF						IVM					
		loc.			non loc.			loc.			non loc.		
		s.	tr. s.	pix.	s.	tr. s.	pix.	s.	tr. s.	pix.	s.	tr. s.	pix.
1	9	7	0	50.9	3.7	1.7	49.4	13	7.8	55	10.2	8	51.3
2	18	5.9	0.2	54	4.3	2.7	47.8	16.3	10.7	79.3	12.7	11.2	71.7
3	17	7.2	0	48	4	1	46.3	14.9	14.3	60.9	12.7	12.1	58.2
4	39	6.8	0	50.8	3.1	1.9	48.4	15	10.2	63.5	11.2	9.7	59.9
5	59	4.4	0	50.8	3.4	0.7	47.3	15.3	7	63.3	11.1	6.6	62.5
6	74	62.5	0.2	113.4	50.7	5.1	98.9	80.3	55.1	125.6	66.4	47.4	115.2
7	106	6.1	0	54.7	3.5	1.2	53.8	17.8	9.2	78.1	11.7	8	66.3
8	78	4.8	0	58	4.7	0	52.1	18.9	9.7	83.7	14.6	10.2	73
9	78	5.8	0	57.6	3	0.7	52.3	19.2	10.7	81.9	11.9	7.5	74.3
10	96	13.6	0	79.7	8.6	2.4	71.8	37.9	26.2	108.4	24.2	34.5	96.7
11	110	4.6	0	51.8	3.8	0.2	56.3	14.5	7.5	71	8.9	7.5	62.1

Table 4. Relative values (in %) of wrong assignments for the dataset *Münster*. For each setting, same results as in Table 2 are reported

feat. set	num. of feat.	RF						IVM					
		loc.			non loc.			loc.			non loc.		
		s.	tr. s.	pix.	s.	tr. s.	pix.	s.	tr. s.	pix.	s.	tr. s.	pix.
1	8	5.6	0.1	15.1	1.9	2.3	9.6	39.4	129.1	52.6	33.2	111.4	47.6
2	16	5.6	0.1	14.9	1.8	2.5	9.4	28.4	88.2	39.1	17.2	58.3	30.1
3	16	4.6	0	12	1.7	1.1	8.6	21.1	64.7	32.6	16.4	52.6	28.7
4	38	3.3	0	12.5	1.8	1.9	9.2	9.3	22.8	20.2	4.3	10.5	14.2
5	58	1.7	0	10.4	1	0.1	8.4	7.4	10.1	19	5.4	6.6	16.1
6	72	14.9	0	30.2	9.5	2.7	24.3	32.3	49.8	50.1	21.7	38.1	38.9
7	104	2.2	0	12	1.5	0.0	10.1	12.9	23.4	29.8	7.6	14.2	19.8
8	76	2.6	0	13.1	1.8	0.3	10.3	19.6	37.7	34.9	12.4	20.3	24.6
9	76	3.6	0	14.2	2.2	0.6	10.5	14.9	37.1	27.7	6.5	17.4	17.4
10	96	5.3	0	20.2	5.0	0.8	17.8	15.9	24.7	33.8	10.4	15.2	23.9
11	108	2.5	0	12.3	1.3	0.2	9.6	17.6	37.2	31.8	11.2	24.3	22.9

1. core features: Superpixel-wise mean values of relative elevation, planarity, saturation and unfiltered channels ■
 2. core features with variances
 3. core features extended by those based on stripes
 4. core features plus filter banks ■
 5. no variance-based features ■
 6. only image-based features
 7. no features based on stripes ■
 8. all features without filter banks
 9. all features without morphological profiles ■
 10. all features not based on means and variances of the unfiltered image channels
 11. all features ■
- One can see that we start with a relatively small feature set, to which we add some feature groups, and we end at the full feature set preceded by those, which can be described as difference set between the full feature set and a certain group of features. The motivation to consider difference sets is given by the so-called *xor*-problem: If we have two features useless on its own and useful together, we cannot reach an

improvement in the result by adding just one of them. However, we notice degrading performance while considering difference sets. The values for relative misclassifications are reported in Tables 2, 3, and 4 for all feature sets specified above while the graphical dependencies on α are shown in Figures 4, 5, and 6. For design-technical reasons, we show only the most characteristic and promising curves, which are specified by colored squares in the enumeration above and in the tables. Additionally, in order not to go too much beyond the scope of this work, we will not track the qualitative assessment of the performance with respect to changing α , feature set, and classifier, but restrict ourselves, instead, to some general remarks.

Observations

We start our observations by considering both classifiers for which we refer to Tables 2, 3, and 4 and Figures 4, 5, and 6. One can see that the Random Forest classifier has a significantly lower error rate for all feature choices for all datasets. With respect to validation of training data, these numbers are almost zero for the RF classifier and can achieve some double-digit percentages for IVM. The computation time for this latter classifier is considerably larger also.

Our next question, namely, how the performance is influenced by the non-local optimization, is worth discussing from the point of view of overall performance and choice of smoothness parameter α with respect to the feature set. As one can easily expect, there is always a more or less significant improvement of the performance with increasing α and after some point, it becomes worse since too many superpixels are oversmoothed. The more features are considered, the closer to zero is the aforementioned point: It happens because for large feature sets, the decisions taken by the classifier are less categorical, the unary potentials, or the data term in Equation 6 are more diluted, and smaller values of α are sufficient to change labels. Besides, we can see that α leading to the best performance is higher for IVM than for RF, which corroborates our assumption that this last classifier is less categorical and more flexible. According to Niculescu-Mizil and Caruana (2005), RF pulls the output probabilities towards the middle of the interval (0; 1) while IVM, not analyzed there, turns out to pull them towards the margins. Another important and intuitively clear conclusion is that overall performance is improved at the cost of superpixels used for training. In the question how far down we may arrive, two conclusions may be drawn. First, it seems that curves barely intersect, that is, a good local result (low point at the beginning) usually has a consequence that the whole curve is quite low. However, for example, in Figure 6, left, the green, red and blue curves are very shallow while the decay of the other curves is among the most considerable. This means that the benefit of non-local optimization is low for those feature sets with a high percentage of features already incorporating neighbors (filter banks, morphological profiles and stripes). The less influence of filtered and stripe-based features is, the better is the performance of non-local optimization.

The question about the influence of the feature set seems to be less trivial because it varies with every dataset and from one classifier to another.

Generally, 3D data tend to be very important, as expected. The same statement can be made for unfiltered channels since their number is small and, at the same time, their absence degrades the performance significantly. Stripes bring improvements where the third row of Tables 2, 3, and 4 has lower error rates than the first row and the eleventh row has lower error rates than the seventh row. One can see that for Random Forest classifier, features based on stripes improve the performance for the *Vaihingen* dataset, despite the streets are predominantly curvy, and partly for the remaining datasets. Unfortunately, in case of IVM classifier, these features only were useful for the *Münster* dataset.

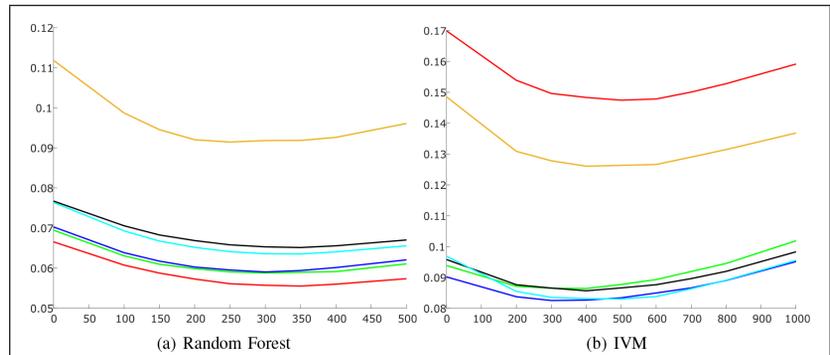


Figure 4. Results for the non-local optimization for the chosen feature sets, dataset *Vaihingen*: Superpixel-wise relative numbers of misclassifications as a function of smoothness parameter α in Equation 6. For technical reasons (integer programming), the negative likelihoods were multiplied by 2048 ($=2^{12}$). Feature sets corresponding to the colors are indicated in the tables. On the left, RF and on the right, IVM.

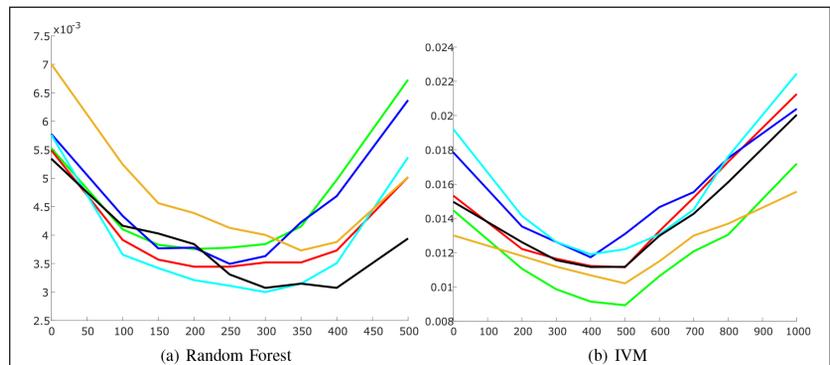


Figure 5. Results for the non-local optimization for the chosen feature sets, dataset *Munich*. For further explanations, please refer to the caption of Figure 4.

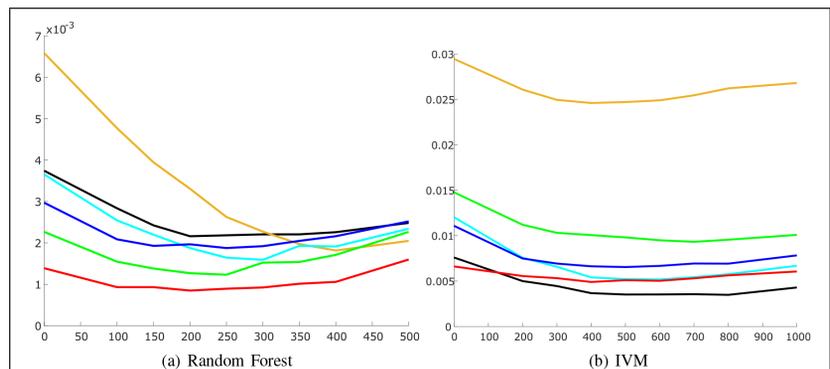


Figure 6. Results for the non-local optimization for the chosen feature sets, dataset *Münster*. For further explanations, please refer to the caption of Figure 4.

Overall, because of its robustness against redundant and irrelevant features, large feature sets do in general improve the performance of the Random Forest classifier. From the full set, one can discard stripes, filter bank, morphological profiles, or unfiltered channels and will obtain comparable or slightly worse results. Analogically, to the core feature set, one can add variances, responses of the filter bank, stripes, etc., and the results mostly become better. The only exception is involving variances in the *Münster* dataset: It seems that without them, the number of wrongly classified superpixels is significantly lower for local and non-local algorithms. In other datasets, the superpixels' size is larger, thus, their second moments may incorporate certain valuable information, and thus cannot be assumed as useless. For IVM, the situation is less clear. Feature sets with a successful performance of Random Forest are not necessarily so for our second classifier, which is not further surprising since different learners also differ by their mechanisms and ability to treat combinations of features. Besides, feature sets that work quite well for the *Münster* dataset seem to fail for *Vaihingen* dataset and vice versa. The configuration *core features extended by filter banks* (black curves) yields relatively good results in all three datasets. Most likely, since IVM does not perform well for many redundant features, a conservatively chosen set of unfiltered and filtered channels of the image and elevation is the best choice to make.

Concerning the correlation of pixel-wise and superpixel-wise accuracies as a function of superpixel size, one can easily estimate the probability P_{pix} of a pixel to be assigned to a wrong class using our evaluation metrics:

$$P_{\text{pix}} = \frac{sn^2 + bn\beta}{Sn^2} \approx P_{\text{segm}} + n\beta P_{\text{border pix}}, \quad (7)$$

where P_{segm} is the superpixel-wise probability of incorrect assignment, $P_{\text{border pix}}$ is that for border pixels (more under-segmentations for complicated borders expected), $0 \leq \beta \leq 1$, because such a border pixel may or may not be classified correctly, b is the number of border pixels, and S and s are total numbers of segments and that of incorrectly assigned ones, respectively. In Equation 7, we made two self-explanatory assumptions that the total number of pixels is roughly Sn^2 while for a good superpixel algorithm, the number of labeled superpixels almost equals S though it can be bounded by $S - b/n$. The second term of Equation 7 is caused by under-segmentations and thus, it becomes clear that the pixel-wise result typically degrades with a growing superpixel size. However, the first term is becoming smaller with the growing superpixel size since a superpixel of a larger area bundles more information, analogously to a larger receptive field in a deep learning algorithm. These theoretical considerations were confirmed by our experiments on *Vaihingen* dataset: A growing superpixel size makes the results to become worse on the pixel and slightly better on the superpixel level. The quantitative evaluation will be an issue of our future work, in which comparisons of a superpixel-based approach with that based on CNNs are currently being made.

There are two reasons why the results for the *Vaihingen* data are worse compared to the other two datasets: First, the quality of the image (8-bit images, 3 channels) and elevation data (3D stereo from optical images) is worse. Second, the training data extracted from freely available sources in the methodology section for the other two datasets are "easier" to classify than the elaborately crafted masks for the *Vaihingen* data. In other words, many complicated situations remained not considered during training data acquisition (errors of the first kind). Finally, we present some qualitative results for the dataset *Vaihingen* (Figures 7 and 8), *Munich* (Figure

9) and *Münster* (Figure 10). From these illustrations, we can see that while the objects' interiors are mostly clean, misclassifications overwhelmingly occur at the boundaries between classes. To the major part, this happens because of under-segmentations, which is reflected by higher percentage of pixel-based than superpixel-based errors in the Tables 2, 3, and 4 and in the off-diagonal elements of the confusion matrices. Though we do not present detailed results for the numbers of confusions between classes, they provide the evidence to another observation one can make from Figures 7 through 10: Most misclassifications happen between the classes *street* and *grass* in shadowy regions, followed by *tree* and *grass*. This is because segment-wise assignment of shrubbery to the tree or grass class is not a trivial issue in aerial images. Additionally, in the dataset *Vaihingen* the ground truth is not always correct; note that the construction highlighted by question mark in Figure 7, top left, can hardly be grass since its height is above 2.5 m. On the positive side, it is surprising that in the dataset *Munich* the classifier was able to assign the correct class (almost everywhere) to the shadowy street in Figure 9, top two rows, and to the building with the grass-covered roof (two bottom rows), even though only a few training examples are extracted from there. Sadly, difficulties are noticed in the park where many leafless trees are confused with roads. The dataset *Münster* is probably the easiest since superpixels close to the borders of regions are seldomly labeled because of the low resolution and were therefore excluded from evaluation. Since only building borders are captured cleanly, the algorithm assigns the class *building* to the areas of shadowy forest borders. Partly, these insufficiencies can be corrected by the non-local optimization. However, near the highway bridge of Figure 10, two bottom rows, it seems to bring more harm than good. Overall, non-local optimization suppresses high-frequency noise, and thus plays a positive part, especially taking into account the almost negligible computing time. Higher-order potentials may bring an additional improvement, however at the cost of many parameters to be chosen by the developer. Despite interesting mathematical concepts underlying these potentials (e.g., Kohli *et al.*, 2009), in the authors' opinion, the key strategy for an improved performance is to foster development of high-level features incorporating context information and allowing to overcome borders of superpixels. For example, CNNs already use quite large receptive fields thus allowing to learn complex environments without a strong necessity of non-local optimization.

Conclusions and Future Work

In this paper, we presented a workflow for classification of aerial images. The procedure is supported by OSM data, which is thoroughly prepared using a fast segmentation method and the most distinctive features, like relative elevation and NDVI. For classification, whereby four classes were considered, around 100 features are chosen. The classification takes part either pixel- or superpixel-wise and can be performed by any probabilistic classifier since the output probabilities are evidently smoothed by a non-local optimization algorithm, such as graph-cuts. We considered here Random Forest and Import Vector Machine; however, since according to Niculescu-Mizil and Caruana (2005) it is possible to assign probabilities to output of other state-of-the-art classifiers, the classification tool can be replaced by a more suitable one, if necessary.

Three other important contributions of our workflow are: Incorporation of the possibly incomplete or obsolete OSM data, universal framework for performing operations on segments and superpixels, as well as context-based high-level features supposed to improve the classification in shadowy regions. We performed evaluation using three datasets that

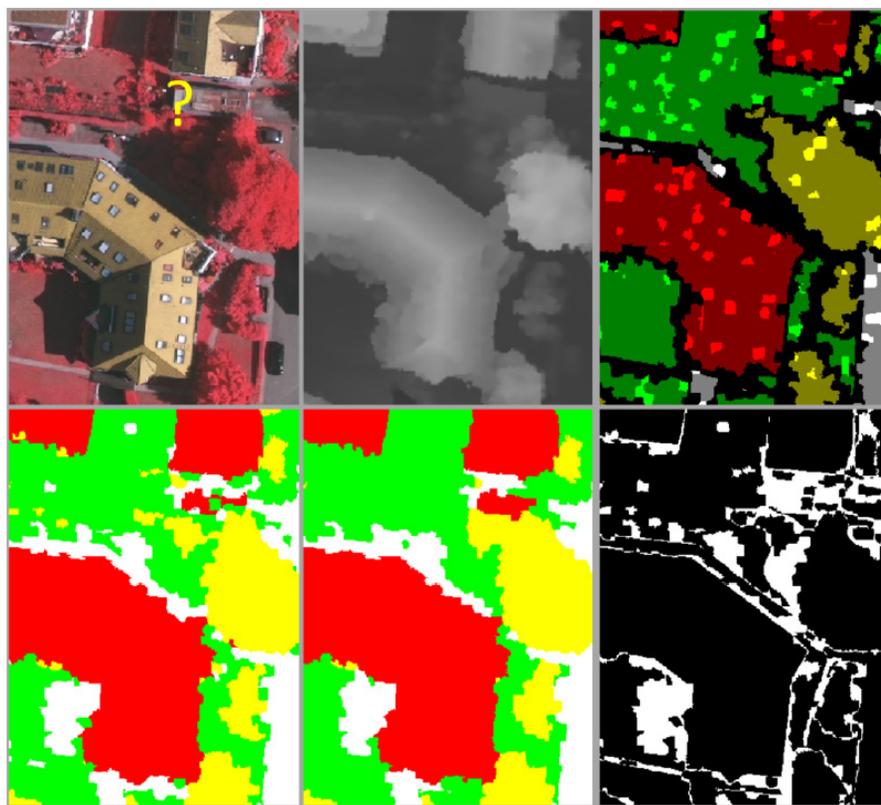


Figure 7. Input data and qualitative results for the dataset *Vaihingen*, fragment from patch 1. Top row, left to right: Optical data, elevation data, labeled and training data (highlighted). Bottom row, left and middle: Superpixel-wise classification result of Random Forest without and with non-local optimization, respectively; right, pixel-wise evaluation result of the configuration in the middle. For more details, see caption of Figure 3.

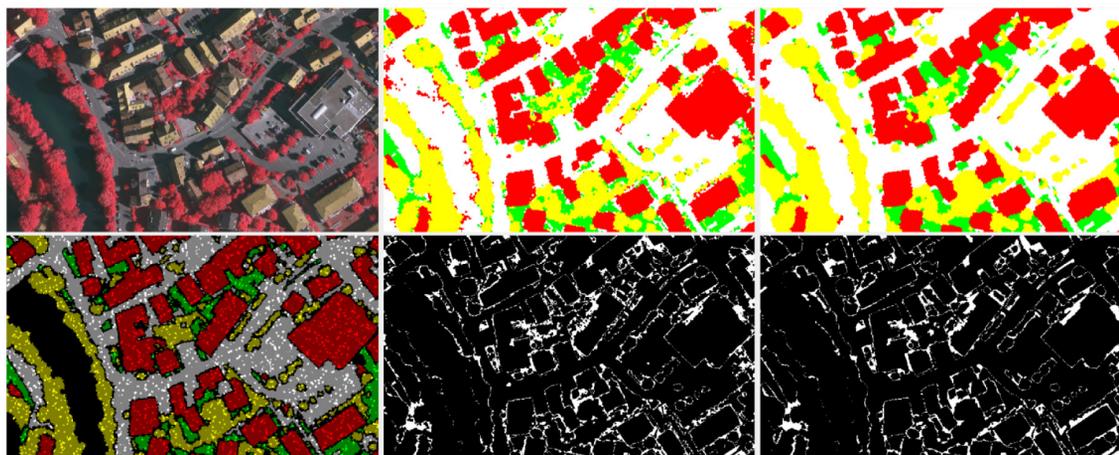


Figure 8. Qualitative results for the dataset *Vaihingen*, patch 11. Top row, left to right: Optical data, superpixel-wise classification result of Random Forest without and with non-local optimization, respectively. Bottom: Labeled and training data (highlighted), pixel-wise evaluation result for the configurations of the top row. For more details, see caption of Figure 3.

show different scenes and exhibit different challenges. Overall encouraging results of only 5.6% wrongly classified superpixels have been obtained for the ISPRS benchmark dataset *Vaihingen* while for both remaining datasets, plausible qualitative results, even in challenging situations, could be reported. On the pixel level, the best result of around 87% correctly classified pixels (for the benchmark dataset) is below the state of the art. However, one should take into account that we (1) performed classification on superpixels (in other

words, any under-segmentation results in errors on the pixel level); (2) did not delete (as many authors do) pixels close to the border of classes; (3) used only some 10% of the whole arsenal of labeled data for learning; and (4) only imposed a primitive smoothness prior, in which unequal labels for pairs of neighboring superpixels are penalized independently on these labels. Introducing smarter filters (Schindler, 2012), context information, e.g., minimum size of a hypothesized building, or topological constraints on road systems (Wegner

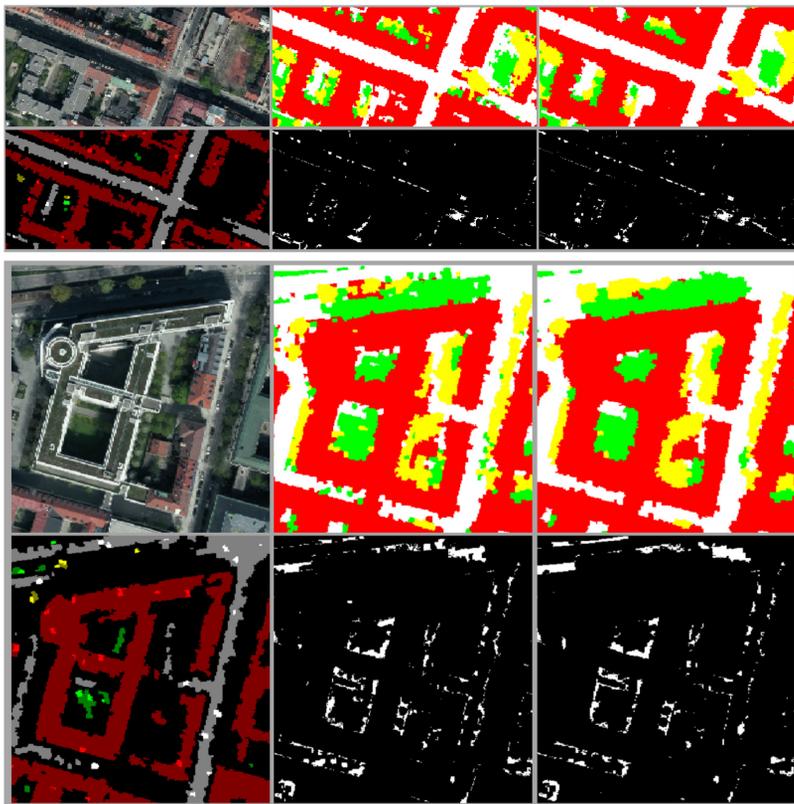


Figure 9. Qualitative results for the dataset *Munich*. Top two rows: Around a shadowy street, bottom two rows: Around a building with a roof covered by grass. For the meaning of subimages, please refer to Figure 8.

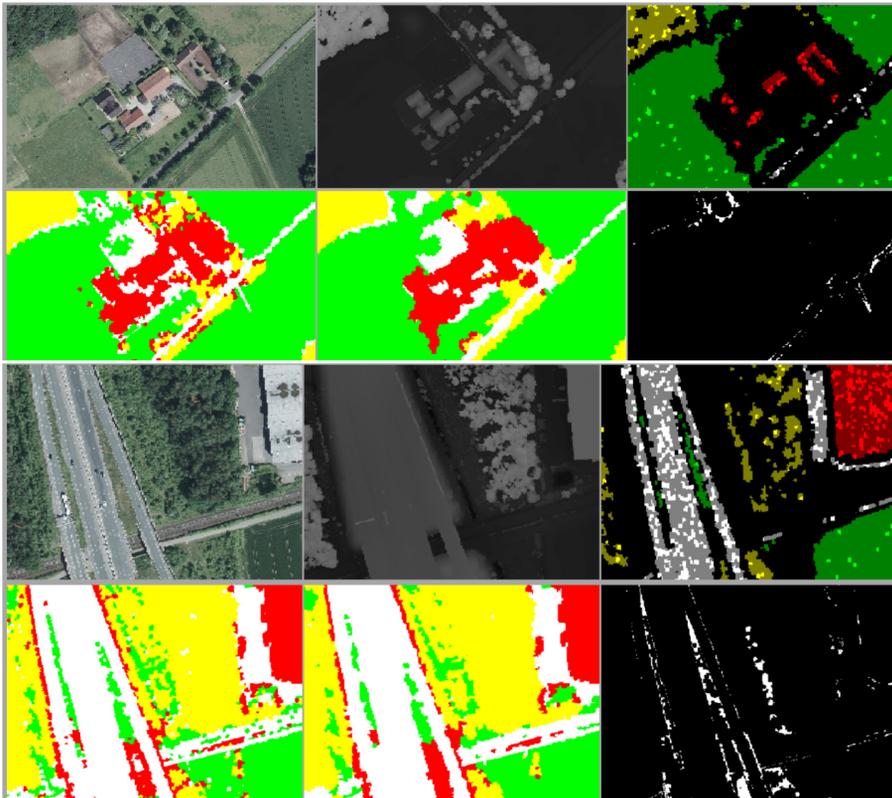


Figure 10. Qualitative results for the dataset *Münster*. Top two rows: Around a farm, bottom two rows: Highway bridge. For the meaning of sub-images, please refer to Figure 7.

et al., 2015), would improve the results at cost of an increased number of parameters. Since for this paper, we were interested in a general algorithm, these modifications and other post-processing steps will be postponed for future work.

Regarding the difference in performance for both classifiers, the main conclusion can be drawn that Random Forest outperforms Import Vector Machine. Moreover, our experiments indicate that the feature set for Random Forest can be tendentiously quite large (many features from different sources) and then smoothed by a moderate constant. If processed superpixel-wise, it is sufficient to consider mean values and exclude the variances of features, especially for small superpixel sizes. For Import Vector Machine, we could see that the feature sets producing best results should not contain too many features and this learner reacts on completely different feature sets as Random Forest.

Lastly, we wish to name two aspects which were not included in this contribution, but are nevertheless very important and represent the main focus of our further work. As we could see in Training Data Section, by means of our widely automatized module for training data extraction, huge amounts on data can be collected and exploited for learning. However, this data is easy to classify which certainly explains the fact that *Munich* and *Münster* data exhibit far better results than the benchmark dataset *Vaihingen*. Additionally, to train the classifier, the whole data was arbitrarily reduced and balanced, even despite the fact that we actually know which fragments of data are easy and which are difficult to assign. To perform a more sophisticated selection of training examples, some work must be done to reduce the errors of the first kind during training data selection and maybe even to interpose an active learning method, such as (Tuia *et al.*, 2012). Hence, the first direction of future work will be to investigate how to choose – quantitatively and qualitatively – training data to improve the algorithm performance. Secondly, since considerable amount of training data is available, we shall investigate to what extent approaches based on CNNs can be applied to our superpixel-based algorithm.

Acknowledgments

The authors thank Dr. Mathias Rothermel for providing the input data for the *Munich* dataset as well as Prof. Ribana Roscher for providing background information about Important Vector Machines and reading parts of the manuscript.

References

- Badrinarayanan, V., A. Handa, and R. Cipolla- Segnet 2015. A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling, *arXiv preprint arXiv:1505.07293*.
- Bellens, R., S. Gautama, L. Martinez-Fonte, W. Philips, J. C.-W. Chan, and F. Canters, 2008. Improved classification of VHR images of urban areas using directional morphological profiles *IEEE Transactions on Geoscience and Remote Sensing*, 46(10):2803–2813.
- Biau, G., Analysis of a random forests model, 2012. *Journal of Machine Learning Research*, 13(April):1063–1095.
- Boykov, Y., O. Veksler, and R. Zabih, 2001. Fast approximate energy minimization via graph cuts, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222– 1239.
- Breiman, L. Random forests, *Machine Learning*, 45(1):5–32, 2001.
- Bruzzone, L., and L. Carlin, 2006. A multilevel context-based system for classification of very high spatial resolution images, *IEEE Transactions on Geoscience and Remote Sensing*, 44(9):2587–2600.
- Bulatov, D. and S. Warnke, 2017. Extraction of road pixels from airborne image and elevation data: Focusing on feature selection, *Proceedings of the IEEE International Conference on Digital Image Computing: Techniques and Applications*, pp. 1–7.
- Bulatov, D., G. Häufel, J. Meidow, M. Pohl, P. Solbrig, and P. Werner, 2014. Context-based automatic reconstruction and texturing of 3D urban terrain for quick-response tasks, *ISPRS Journal of Photogrammetry and Remote Sensing*, 93:157–170.
- Bulatov, D., G. Häufel, and M. Pohl, 2016a. Vectorization of road data extracted from aerial and UAV imagery, *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pp. 567–574.
- Bulatov, D., I. Wayand, and H. Schilling, 2016b. Automatic tree-crown detection in challenging scenarios. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 41:575–582.
- Burns, J.B., A.R. Hanson, and E.M. Riseman, 1986. Extracting straight lines. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (4):425–455.
- Butenuth, M. and C. Heipke, 2012. Network snakes: Graph-based object delineation with active contour models, *Machine Vision and Applications*, 2012. 23(1):91–109.
- Cula, O.G., and K.J. Dana, 2004. 3D texture recognition using bidirectional feature histograms, *International Journal of Computer Vision*, 59(1):33–60.
- Dalal, N., and B. Triggs. 2005. Histograms of oriented gradients for human detection, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Volume 1, pp. 886–893.
- DeLong, A. Osokin, H.N. Isack, and Y. Boykov, 2012. Fast approximate energy minimization with label costs, *International Journal of Computer Vision*, 96(1):1–27.
- Edelsbrunner, H. and E.P. Mücke, 1994. Three-dimensional alpha shapes, *ACM Transactions on Graphics (TOG)*, 13(1):43–72.
- Fritz, S., I. McCallum, C. Schill, C. Perger, R. Grillmayer, F. Achard, F. Kraxner, and M. Obersteiner, 2009, Geo-Wiki.Org: The use of crowdsourcing to improve global land cover, *Remote Sensing*, 1(3):345–354.
- Fröhlich, B., E. Bach, I. Walde, S. Hese, C. Schmullius, and J. Denzler, 2013. Land cover classification of satellite images using contextual information, *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 3:W1.
- Fulkerson, B., A. Vedaldi, and S. Soatto, 2009. Class segmentation and object localization with superpixel neighborhoods, *Proceedings of the IEEE International Conference on Computer Vision*, pp. 670–677.
- Genuer, R., J.-M. Poggi, and C. Tuleau-Malot, 2010. Variable selection using random forests, *Pattern Recognition Letters*, 31(14):2225–2236.
- Gerke, M., and J. Xiao, 2014. Fusion of airborne laserscanning point clouds and images for supervised and unsupervised scene classification, *ISPRS Journal of Photogrammetry and Remote Sensing*, 87:78–92.
- Geusebroek, J.-M., R. Van den Boomgaard, A.W.M. Smeulders, and H. Geerts, 2001. Color invariance, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(12):1338–1350.
- Gross, H., and U. Thönnessen, 2006. Extraction of lines from laser point clouds, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36(Part 3/W49):86–91.
- Guo, L., N. Chehata, C. Mallet, and S. Boukir, 2011. Relevance of airborne lidar and multispectral image data for urban scene classification using random forests, *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(1):56–66.
- Haala, N., 2005., Multi-Sensor-Photogrammetrie – Vision oder Wirklichkeit? Habilitation, Deutsche Geodatische Kommission, Munchen (in German).
- Haala, N., and M. Rothermel, 2012. Dense multi-stereo matching for high quality digital elevation models, *Photogrammetrie-Fernerkundung-Geoinformation*,(4):331–343.
- Haklay, M., 2010. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets, *Environment and Planning B: Planning and design*, 37(4):682–703.
- Heipke, C., 2010. Crowdsourcing geospatial data, *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(6):550–557.
- Hirschmüller, H., 2008. Stereo processing by semi-global matching and mutual information, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2):328–341.
- Huang, X., and L. Zhang, 2013. An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery, *IEEE Transactions on Geoscience and Remote Sensing*, 51(1):257–272.
- Huang, X., Q. Lu, and L. Zhang, 2014. A multi-index learning approach for classification of high-resolution remotely sensed images over urban areas, *ISPRS Journal of Photogrammetry and Remote Sensing*, 90:36–48.
- Huang, X., C. Weng, Q. Lu, T. Feng, and L. Zhang, 2015. Automatic labelling and selection of training samples for high-resolution remote sensing image classification over urban areas, *Remote Sensing*, 7(12):16024–16044.
- Huang, X., D. Wen, J. Li, and R. Qin, 2017. Multi-level monitoring of subtle urban changes for the megacities of China using high-resolution multi-view satellite imagery, *Remote Sensing of Environment*, 196:56–75.
- Kaiser, P., J.D. Wegner, A. Lucchi, M. Jaggi, T. Hofmann, and K. Schindler, 2017. Learning aerial image segmentation from online maps, *IEEE Transactions on Geoscience and Remote Sensing*, 55(11):6054–6068.
- Kluckner, S., T. Mauthner, P.M. Roth, and H. Bischof, 2009. Semantic classification in aerial imagery by integrating appearance and height information, *Proceedings of the Asian Conference on Computer Vision*, Springer, pp. 477–488.
- Kohli, P., L. Ladicky, and P.H.S. Torr, 2009. Robust higher order potentials for enforcing label consistency, *International Journal of Computer Vision*, 82(3):302–324.
- Lafarge, F., and C. Mallet, 2012. Creating large-scale city models from 3D-point clouds: A robust approach with hybrid representation, *International Journal of Computer Vision*, 99(1):69–85.
- Long, J., E. Shelhamer, and T. Darrell, 2015. Fully convolutional networks for semantic segmentation, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440.
- Lopes, P., C. Fonte, L. See, and B. Bechtel, 2017. Using OpenStreetMap data to assist in the creation of LCZ maps, *Proceedings of Urban Remote Sensing Event (JURSE), 2017 Joint, IEEE*, pp. 1–4
- Maas, A., F. Rottensteiner, and C. Heipke, 2017, Classification under label noise based on outdated maps, *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*.
- Maggiori, E., Y. Tarabalka, G. Charpiat, and P. Alliez, 2016. High-resolution semantic labeling with convolutional neural networks, *arXiv preprint arXiv:1611.01962*, 2016.

- Mattys, G., S. Wang, S. Fidler, and R. Urtasun, 2015. Enhancing road maps by parsing aerial images around the world, *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1689–1697.
- Mena, J.B. 2006. Automatic vectorization of segmented road networks by geometrical and topological analysis of high resolution binary images, *Knowledge-Based Systems*, 19(8):704–718.
- Mills, G., J. Ching, L. See, B. Bechtel, and M. Foley, 2015. An introduction to the wudapt project, *Proceedings of the 9th International Conference on Urban Climate, Toulouse*.
- Montoya-Zegarra, J.A., J.-D. Wegner, L. Ladicky, and K. Schindler, 2015. Semantic segmentation of aerial images in urban areas with class-specific higher-order cliques, *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2(3):127.
- Niculescu-Mizil, A., and R. Caruana, 2005. Predicting good probabilities with supervised learning, *Proceedings of ICML*.
- Pesaresi, M., and J.A. Benediktsson, 2001. A new approach for the morphological segmentation of high-resolution satellite imagery, *IEEE Transactions on Geoscience and Remote Sensing*, 39(2):309–320.
- Roscher, R., B. Waske, and W. Förstner, 2012. Incremental import vector machines for classifying hyperspectral data, *IEEE Transactions on Geoscience and Remote Sensing*, 50(9):3463–3473.
- Rother, C., 2006. A comparative study of energy minimization methods for Markov Random Fields, *Proceedings of the IEEE European Conference on Computer Vision*, pp.16–29.
- Rumpler, M, A. Irschara, A. Wendel, and H. Bischof, 2012. Rapid 3D city model approximation from publicly available geographic data sources and georeferenced aerial images, *Proceedings of the Computer Vision Winter Workshop (CVWW)*, 2012.
- Schilling, H., D. Bulatov, and W. Middelmann, 2015. Object-based detection of vehicles in airborne data, *Proceedings of Image and Signal Processing for Remote Sensing XXI*, Volume 9643, page 964316, International Society for Optics and Photonics.
- Schilling, H., D. Bulatov, and W. Middelmann, 2018. Object-based detection of vehicles using combined optical and elevation data, *ISPRS Journal of Photogrammetry and Remote Sensing*, 136:85–105.
- Schindler, K, 2012. An overview and comparison of smooth labeling methods for land-cover classification, *IEEE Transactions on Geoscience and Remote Sensing*, 50(11):4534–4545.
- Shackelford, A.K., and C.H. Davis, 2003a, A combined fuzzy pixel-based and object-based approach for classification of high-resolution multispectral data over urban areas, *IEEE Transactions on Geoscience and Remote Sensing*, 41(10):2354–2363.
- Shackelford, A.K., and C.H. Davis, 2003b, A self-supervised approach for fully automated urban land cover classification of high resolution satellite imagery, *Proceedings of the International Symposium on Remote Sensing and Data Fusion Over Urban Areas*, pp. 14–16.
- Soergel, U., E. Cadario, H. Gross, A. Thiele, and U. Thoennessen, 2006. Bridge detection in multiaspect high-resolution interferometric SAR data, *Proceedings of EUSAR 2006*.
- Sohn, G., Y. Jwa, H.B. Kim, and J. Jung 2012. An implicit regularization for 3D building rooftop modeling using airborne LIDAR data, *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2 (3):305–310.
- Spreckels, V., L. Syrek, and A. Schlienckamp 2010. Dgpf-project: Evaluation of digital photogrammetric camera systems – Stereoplotting, *Photogrammetrie-Fernerkundung-Geoinformation*, 2010 (2):117–130.
- Straub, B.-M., 2003. Automatic extraction of trees from aerial images and surface models, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 34(3/W8):157–166.
- Sun, X., X. Lin, S. Shen, and Z. Hu, 2017. High-resolution remote sensing data classification over urban areas using random forest ensemble and fully connected conditional random field, *ISPRS International Journal of Geo-Information*, 6(8):245.
- R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother, 2006. A comparative study of energy minimization methods for markov random fields. *Proceedings of the IEEE European Conference on Computer Vision*, pp.16–29.
- D. Tuia, J. Muñoz-Marí, and G. Camps-Valls, 2012. Remote sensing image segmentation by active queries, *Pattern Recognition*, 45(6):2180–2192.
- Varma, M., and A. Zisserman, 2005. A statistical approach to texture classification from single images, *International Journal of Computer Vision*, 62(1-2):61–81.
- Veksler, O., Y. Boykov, and P. Mehrani, 2010. Superpixels and supervoxels in an energy optimization framework, *Proceedings of the European Conference on Computer Vision*, Springer, pp. 211–224
- Vogt, K., A. Paul, J. Ostermann, F. Rottensteiner, and C. Heipke, 2017. Boosted unsupervised multisource selection for domain adaptation, *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, 4.
- Warnke, S., 2017. *Variable Selection for Random Forest and Logistic Regression applied to Road Network Extraction from Aerial Images*, Master thesis at Karlsruhe Institute of Technology, Germany.
- Wassenberg, J., W. Middelmann, and P. Sanders, 2009. An efficient parallel algorithm for graph-based image segmentation, *Proceedings of the International Conference on Computer Analysis of Images and Patterns*, Springer, pp. 1003–1010.
- Wegner, J.D., J.A. Montoya-Zegarra, and K. Schindler, 2015. Road networks as collections of minimum cost paths, *ISPRS Journal of Photogrammetry and Remote Sensing*, 108:128–137.
- Winn, J., A. Criminisi, and T. Minka, Object categorization by learned universal visual dictionary, *Proceedings of the IEEE International Conference on Computer Vision*, Volume 2, pp 1800–1807.
- Yoo, H.Y., K. Lee, and B.-D. Kwon, 2009. Quantitative indices based on 3D discrete wavelet transform for urban complexity estimation using remotely sensed imagery, *International Journal of Remote Sensing*, 30(23):6219–6239.
- Yuan, J., and A.M. Cheryadat, 2013. Road segmentation in aerial images by exploiting road vector data, *Proceedings of the International Conference on Computing for Geospatial Research and Application (COM. Geo)*, IEEE pp. 16–23.
- Zhang, Q., R. Qin, X. Huang, Y. Fang, and L. Liu, 2015. Classification of ultra-high resolution orthophotos combined with dsm using a dual morphological top hat profile, *Remote Sensing*, 7(12):16422–16440.
- Zhu, J., and T. Hastie, 2005. Kernel logistic regression and the import vector machine, *Journal of Computational and Graphical Statistics*, 14(1):185–205.

SUSTAINING MEMBERS

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.terraremove.com
Member Since: 10/2016

The Airborne Sensing Corporation

Toronto, Canada
www.airsensing.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

Improved Camera Distortion Correction and Depth Estimation for Lenslet Light Field Camera

Changkun Yang, Zhaoqin Liu, Kaichang Di, Yexin Wang, and Man Peng

Abstract

A light field camera can capture both radiance and angular information, providing a novel solution for depth estimation. The paper proposes two improved methods including distortion model optimization and depth estimation refinement for a lenslet light field camera. For distortion model optimization, a novel 14-parameter distortion model that involves sub-aperture images generation is applied to correct the light field camera images. For depth estimation refinement, an algorithm reducing the high influence of outliers on depth estimation in weak texture regions is proposed based on multi-view stereo matching using the cost volume. Experimental results show the projection error has decreased by about 30% and the depth root-mean-squared error on real world images has decreased by about 42% with our distortion correction method and depth estimation method compared with state of art algorithms. It verifies the correctness and effectiveness of our proposed methods and show significant improvement on accuracy of depth map estimation.

Introduction

Light field cameras have become popular in recent years in computational photography, computer vision, and the close range photogrammetry field because they can capture both the radiance and angular information in a single snapshot thanks to a micro-lens array placed between the main lens and sensor. Typical applications include industrial measurement (Heinze *et al.*, 2016), measurement of the growth of plants and animals (Apelt *et al.*, 2015), visual odometry (Dansereau *et al.*, 2011), simultaneous localization and mapping (SLAM) (Dong *et al.*, 2013). Light field cameras can be divided into two categories depending on the distance between the micro-lens array and sensor. In the first category called unfocused plenoptic cameras (Adelson and Wang, 1992; Ng *et al.* 2005), the distance is fixed to be equal to the micro-lens focal length, such as in the commercial products Lytro and Lytro Illum (Lytro, 2017). In the second one called focused plenoptic cameras (Lumsdaine and Georgiev, 2009; Perwaß and Wietzke, 2012), the distance can be changed, such as Raytrix cameras (Raytrix, 2017). In this study, we focus on unfocused plenoptic cameras and use a Lytro Illum camera.

Depth estimation is one of the most important research topics for light field camera image postprocessing. The light field images can be processed for multiple images from different views of the scene, namely sub-aperture images (which will be described in detail in another section). The depth

estimation is based on the disparities observed in the adjacent sub-aperture images, similar to stereo camera approaches. Camera calibration is a necessary prerequisite for accurate depth estimation. A number of methods have been proposed. For the unfocused plenoptic cameras, Dansereau *et al.* (2013) proposed a decoding, calibration, and rectification approach for lenslet light field cameras, in which a 15-parameter camera model was presented for calibration and distortion correction. Cho *et al.* (2013) calibrated a light field camera by searching for local maximization and estimating the rotation of the micro-lens array in the frequency domain based on Dansereau *et al.* (2013). Bok *et al.* (2014) proposed a more accurate calibration method for a micro-lens light field camera based on line features extracted from raw images directly. However, in these calibration algorithms, distortion corrections are all based on the radial distortion model, which does not fit well with the lenslet light field camera. The unfocused plenoptic camera calibration remains to be an important yet challenging task for precision improvement of the subsequent depth estimation. For focused plenoptic cameras, some methods on the metric calibration have been proposed (Heinze *et al.*, 2016; Zeller *et al.*, 2016; Strobl and Lingenauber, 2016), which are beyond the scope of this paper and will not be detailed in the following sections.

Recently a number of depth estimation algorithms for light field images have been proposed. Yu *et al.* (2013) explored the 3D geometry of line in a light field image and derived a disparity map using line matching between sub-aperture images. Wanner and Goldluecke (2013) proposed a local depth estimation algorithm using a structure tensor to compute local slopes in epipolar plane image (EPI). Tao *et al.* (2013) proposed a fusion approach that combined defocus and correspondence cues to estimate the scene depth using EPI, and the global smoothness of depth map was refined by Markov random fields. Tasic *et al.* (2014) proposed a depth estimation algorithm by defining a description of EPI texture and mapping this texture to scale-depth space. Sabater *et al.* (2014) proposed a depth estimation algorithm based on block-matching using the sub-aperture images without demosaicking. Compared with above algorithms, Jeon *et al.* (2015) and Zhang *et al.* (2015) achieved the sub-pixel shifts estimation of sub-aperture images using the phase shift theorem in the Fourier domain to obtain an accurate disparity map. In addition, Kim *et al.* (2013) estimated disparity maps using the 4D light field captured by a digital single lens reflex (DSLR) with movement. Chen *et al.* (2014) introduced a bilateral consistency metric on the surface camera to estimate stereo matches in the light image in the presence of occlusion. However, the baseline of the light field images used in Kim *et al.* (2013) and Chen *et al.* (2014) are much larger than the baseline of

Changkun Yang is with the State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences; and also the University of Chinese Academy of Sciences, Beijing 100049, China.

Zhaoqin Liu, Kaichang Di, Yexin Wang, and Man Peng are with the State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, No. 20A, Datun Road, Chaoyang District, Beijing 100101, China (liuzq@radi.ac.cn).

Photogrammetric Engineering & Remote Sensing
Vol. 85, No. 3, March 2018, pp. 197–208.
0099-1112/18/197–208

© 2018 American Society for Photogrammetry
and Remote Sensing
doi: 10.14358/PERS.85.3.197

the light field images captured by a lenslet light field camera; thus, their methods are not directly applicable for the lenslet light field camera.

In this paper, two improved methods including distortion model optimization and depth estimation refinement for a lenslet light field camera are proposed so as to achieve higher depth estimation accuracy. First, a novel 14-parameter optical distortion model is applied to correct the aberrations of the lenslet light field camera. Then, a depth estimation algorithm inspired by Jeon *et al.* (2015) is proposed to reduce the influence of outliers on the depth estimation for the light field images. The proposed depth estimation algorithm makes use of the phase shift theorem in the Fourier domain to estimate the sub-pixel shifts of sub-aperture images with an extremely narrow baseline as Jeon *et al.* (2015). Compared with Jeon *et al.* (2015), our depth estimation algorithm divides the center sub-aperture image into strong texture regions and weak texture regions by a strong texture confidence measure. Only costs in strong texture regions are calculated and refined by taking the center sub-aperture image as the guiding image to filter every cost slice to alleviate the coarsely scattered unreliable matches (He *et al.*, 2013). With the cost volume, the multi-label optimization propagates the depth estimation in strong texture regions to the weak texture regions. At last, the disparity refinement (Yang *et al.*, 2007) is used to recover a non-discrete depth map. To verify the effectiveness and accuracy of the proposed algorithm, qualitative and quantitative evaluations have been performed by comparing with Jeon's algorithm using complex real-world scenes. The experimental results show that the proposed algorithm is effective and of higher accuracy on depth estimation using a lenslet light field image.

Geometric Model of Light Field Camera and Improved Distortion Correction

Camera Model and Three-dimensional Measurement Principles

For the traditional camera, its geometric model is based on the collinearity equation. In contrast to a traditional camera, a light field camera has a micro-lens array placed between the main lens and the CCD array. The projection model of a lenslet light field camera is shown in Figure 1. All rays from an arbitrary point passing through the main lens and the micro-lens intersect the CCD array at multiple points. In Figure 1, the rays from the arbitrary point $P(X_c, Y_c, Z_c)$ in the camera coordinate system penetrate the main lens and intersect at (X, Y, Z) , which are the image coordinates in the camera coordinate system. The micro-lens image center (u_c, v_c) is defined as the intersection of the CCD, and a ray that passes through the image center and the micro-lens center. Moreover, the coordinates of a projected point (u, v) in a micro-lens image centered at (u_c, v_c) . The distance L_m is between the main lens and the micro-lens array, the distance L_c is between the main lens and the CCD array, (c_x, c_y) is the principal point.

The projection equation of the lenslet light field camera proposed by Bok *et al.* (2014) is as follows:

$$\begin{bmatrix} \Delta u \\ \Delta v \end{bmatrix} = \frac{1}{K_1 Z_c + K_2} \begin{bmatrix} f_x X_c - Z_c \dot{u}_c \\ f_y Y_c - Z_c \dot{v}_c \end{bmatrix} \quad (1)$$

where f_x and f_y are the principal distances of the main lens. Moreover, $\Delta u = u - u_c$ and $\Delta v = v - v_c$ are the displacements from the micro-lens image center. Here, $\dot{u}_c = u_c - c_x$ and $\dot{v}_c = v_c - c_y$. Parameters K_1 and K_2 are additional intrinsic parameters calculated as follows (Bok *et al.*, 2014).

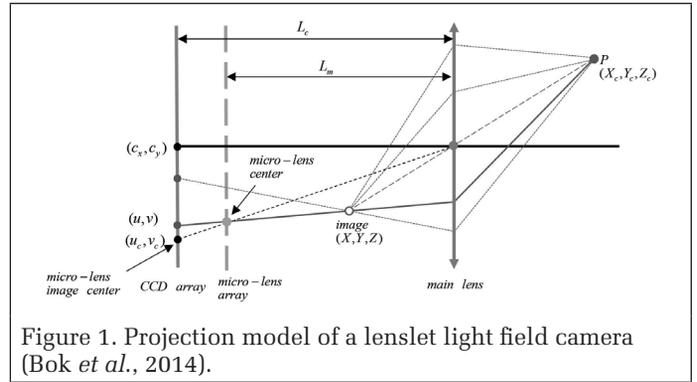


Figure 1. Projection model of a lenslet light field camera (Bok *et al.*, 2014).

$$K_1 \equiv -\frac{(L_m + F)L_c}{(L_m - L_c)F}, \quad K_2 \equiv \frac{L_m L_c}{L_m - L_c} \quad (2)$$

where F is the focal length of the main lens.

It should be noted that the displacements $\Delta u = 0$ and $\Delta v = 0$ for the rays through the center of lens, substituting $\Delta u = 0$ and $\Delta v = 0$ into Equation 1, it becomes the classical collinearity equation:

$$\begin{bmatrix} \dot{u}_c \\ \dot{v}_c \end{bmatrix} = \begin{bmatrix} f_x \frac{X_c}{Z_c} \\ f_y \frac{X_c}{Z_c} \end{bmatrix}. \quad (3)$$

One projection point in the raw image (sensor image) yields one equation in the form of Equation 1, so at least two corresponding points are needed to calculate their 3D coordinates. There are multiple corresponding points in the raw image that come from the same object point when it is out of focus. Therefore, the 3D coordinates of a target point can be computed by Equation 1.

Using the projection equation (1), the theoretical accuracy of three-dimensional measurement using light field images can be analyzed.

From Equation 1, taking two corresponding points in the u -direction as an example, their projection equations are as follows:

$$\begin{aligned} \Delta u_1 &= \frac{1}{K_1 Z_c + K_2} (f_x X_c - Z_c \dot{u}_{c1}) \\ \Delta u_2 &= \frac{1}{K_1 Z_c + K_2} (f_x X_c - Z_c \dot{u}_{c2}) \end{aligned} \quad (4)$$

From Equation 4, we obtain

$$\dot{u}_{c1} - \dot{u}_{c2} = \frac{(\Delta u_2 - \Delta u_1) K_2}{Z_c} + (\Delta u_2 - \Delta u_1) K_1. \quad (5)$$

We define $p = \dot{u}_{c1} - \dot{u}_{c2}$ and $\Delta U = \Delta u_2 - \Delta u_1$ to shorten Equation 5:

$$p = \frac{\Delta U K_2}{Z_c} + \Delta U K_1 \quad (6)$$

From Equation 6, Z_c can be solved by

$$Z_c = \frac{\Delta U K_2}{p - \Delta U K_1} \quad (7)$$

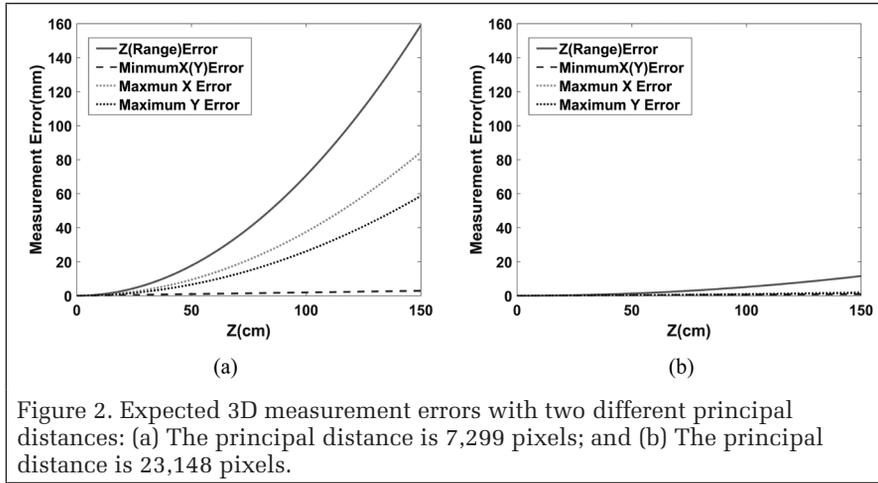


Figure 2. Expected 3D measurement errors with two different principal distances: (a) The principal distance is 7,299 pixels; and (b) The principal distance is 23,148 pixels.

Taking partial derivatives on both sides of Equation 7 and substituting Equation 6 into Equation 7, we get the measurement error in the Z direction as:

$$\sigma_{Z_c} = -\frac{\Delta UK_2}{(p - \Delta UK_1)^2} \sigma_p = -\frac{Z_c^2}{\Delta UK_2} \sigma_p \quad (8)$$

where σ_p is the actual matching error and ΔU is equivalent to the baseline of two sub-aperture images. Equation 8 states that σ_{Z_c} is proportional to Z_c^2 and σ_p but inversely proportional to ΔU . Hence, we can improve the measurement precision in the Z direction by reducing matching error σ_p or increasing ΔU .

From Equation 1, let $\Delta u = \Delta v = 0$. Then, we have:

$$\begin{aligned} X_c &= \frac{\dot{u}_c}{f_x} Z_c \\ Y_c &= \frac{\dot{v}_c}{f_y} Z_c \end{aligned} \quad (9)$$

From Equations 8 and 9, the accuracies (standard errors σ_{X_c} and σ_{Y_c}) of the coordinates in the X and Y directions can be calculated as follows:

$$\begin{aligned} \sigma_{X_c} &= \sqrt{\left(\frac{Z_c^2}{\Delta UK_2}\right)^2 \left(\frac{\dot{u}_c}{f_x}\right)^2 \sigma_p^2 + \left(\frac{Z_c}{f_x}\right)^2 \sigma_{\dot{u}_c}^2} \\ \sigma_{Y_c} &= \sqrt{\left(\frac{Z_c^2}{\Delta UK_2}\right)^2 \left(\frac{\dot{v}_c}{f_y}\right)^2 \sigma_p^2 + \left(\frac{Z_c}{f_y}\right)^2 \sigma_{\dot{v}_c}^2} \end{aligned} \quad (10)$$

where $\sigma_{\dot{u}_c}$ and $\sigma_{\dot{v}_c}$ are the measurement errors of the raw image coordinates in the horizontal and vertical directions, respectively. The measurement errors in X_c and Y_c rely on not only Z_c , but also the position of the object in the raw image. At a given Z_c , the error of X_c is minimized when $\dot{u}_c = 0$, while the largest errors of X_c lie on the left and right margins of the raw image. Similarly, at a given Z_c , the error of Y_c is minimized when $\dot{v}_c = 0$, while the largest errors of Y_c lie on the top and bottom margins of the raw image. Note that the formula for σ_{X_c} , σ_{Y_c} , and σ_{Z_c} is similar to that of the traditional stereo camera (Di and Li, 2007).

Using the standard errors σ_{X_c} , σ_{Y_c} , and σ_{Z_c} , the expected 3D measurement error can be calculated. In order to evaluate the accuracy of depth estimation, we calculated the expected 3D measurement errors for the Lytro Illum camera used in Hazirbas *et al.* (2017), whose principal distance is 7,299 pixels. In

this theoretical analysis, $\Delta U = 1$ pixel, $\sigma_p = 0.14$ pixel, and $\sigma_{\dot{u}_c} = \sigma_{\dot{v}_c} = 14$ pixels, because the size of the raw image is 14 times the size of the sub-aperture image for the Lytro Illum camera. In addition, in order to study the measurement performance in different principal distance, a Lytro Illum camera with the principal distance of 23148 pixels is also analyzed. Figure 2 shows the expected 3D measurement errors for a Lytro Illum camera with the principal distance of 7,299 pixels and the principal distance of 23,148 pixels.

From Equation 8, the expected measurement error σ_{Z_c} is 70.7 mm at 100 cm distance, and σ_{Z_c} is 159.1 mm at 150 cm distance for the Lytro Illum camera with principal distance of 7,299 pixels, as shown in Figure 2a. The expected measurement error σ_{Z_c} is 5.24 mm at 100 cm distance, and σ_{Z_c} is 11.79 mm at 150 cm distance for the Lytro Illum camera with the principal distance of 23,148 pixels, as shown in Figure 2b. Therefore, the expected working range for the light-field camera for measurement purpose in metric range is within 1.5 meters.

Optical Distortion Correction

For the traditional camera, the collinearity equations with additional parameters as radial and decentering distortion is used to calibrate cameras (Brown, 1966; Fryer, 1996). The existing calibration methods (Dansereau *et al.*, 2013; Bok *et al.*, 2014) for the lenslet field camera are also based on the traditional radial distortion model:

$$\begin{bmatrix} \tilde{x} \\ \tilde{y} \end{bmatrix} = \left(1 + k_1 r^2 + k_2 r^4\right) \begin{bmatrix} x \\ y \end{bmatrix} \quad (11)$$

where (\tilde{x}, \tilde{y}) are the undistorted coordinates, (x, y) are the distorted coordinates, and k_1, k_2 are two radial distortion parameters. Because the radial distortion model is based on a pinhole camera model, in which the camera aperture is described as a point and the rays always pass through the point. For a light field camera, the rays that do not pass through the center of the main lens cannot fit well to the model. Therefore, it is not completely suitable to correct the light field images using the radial distortion model. In this paper, a 14-parameter distortion model that considers aberrations is applied to correct distortions for the lenslet light field camera. It needs to be noted that the decentering distortion parameters p_1 and p_2 are not included in our distortion model. There are two reasons: first, the effect of these parameters is very small to the center sub-aperture image of the light field cameras (Lytro Illum) according to our experiments, only about 0.001 pixels or less in the root-mean-squared error (RMSE) of projection errors. Second, the decentering distortion model does not fit well to the rays and do not pass through the center of the main lens.

The flowchart of the distortion correction strategy is shown in Figure 3. It consists of the following steps: (i) solving interior parameters and exterior orientation parameters by light field camera calibration using Bok's calibration method (Bok *et al.*, 2014), as to be described; (ii) computing undistorted checkerboard corners based on the interior parameters and distorted micro-lens image centers; (iii) getting distorted checkerboard corners by matching between the raw image and the center sub-aperture image; (iv) determining 14 initial distortion coefficients using the undistorted and distorted checkerboard corners based on 14-parameter distortion

model; and (v) solving 14 optimized distortion coefficients by recalibration with distorted checkerboard corners, 14 initial distortion coefficients, interior parameters, and exterior orientation parameters.

Light Field Camera Calibration

The lenslet light field camera is firstly calibrated by Bok's method (Bok et al., 2014) to get the interior parameters ($K_1, K_2, f_x, f_y, c_x, c_y, k_1, k_2$) and exterior orientation parameters R, t . The Camera Calibration Toolbox for MATLAB (https://sites.google.com/site/yun-subok/lf_geo_calib) is used to complete the calibration with the raw images of checkerboard pattern captured from the different distances and orientations. The Lytro Illum camera and an example of a checkerboard pattern raw image from Bok's calibration dataset which are available online (Bok, 2017) are shown in Figure 4. The size of the raw image is 7,7285,368 pixels.

The distortion model used in Bok's calibration method (Bok et al., 2014) is the traditional radial distortion model with distortion coefficients k_1 and k_2 . Using the distortion coefficients, the center sub-aperture image to be described can be corrected, which is used to get distorted checkerboard corners by matching with the raw image to be described.

The interior parameters ($K_1, K_2, f_x, f_y, c_x, c_y, k_1, k_2$) and exterior orientation parameters obtained in this subsection are utilized as fixed values to recalibrate camera to get 14 optimized distortion coefficients.

Sub-Aperture Image Generation

A sub-aperture image is a collection of camera rays that pass through the common point on the main lens. Taking the same pixel underneath each micro-lens, the sub-aperture images from different angles can be obtained. As shown in Figure 5a, the green lines, blue lines and yellow lines come from different angles respectively, each of them can make up a sub-aperture image. Figure 5b is an overhead view of Figure 5a.

In this paper, a new method for the sub-aperture images generation is proposed. The micro-lenses are hexagons and arranged like a honeycomb, but the sub-aperture image is an orthogonal grid. Therefore, the center

(red dots in Figure 5b) of the virtual micro-lenses (red dashed box in Figure 5a) are evenly spaced and aligned horizontally and vertically in CCD array. The first center of the virtual micro-lens is at the coordinate (7.5, 7.5) in the CCD array. The space between these centers is the same size as those of the true micro-lenses (14 pixels). Every virtual micro-lens

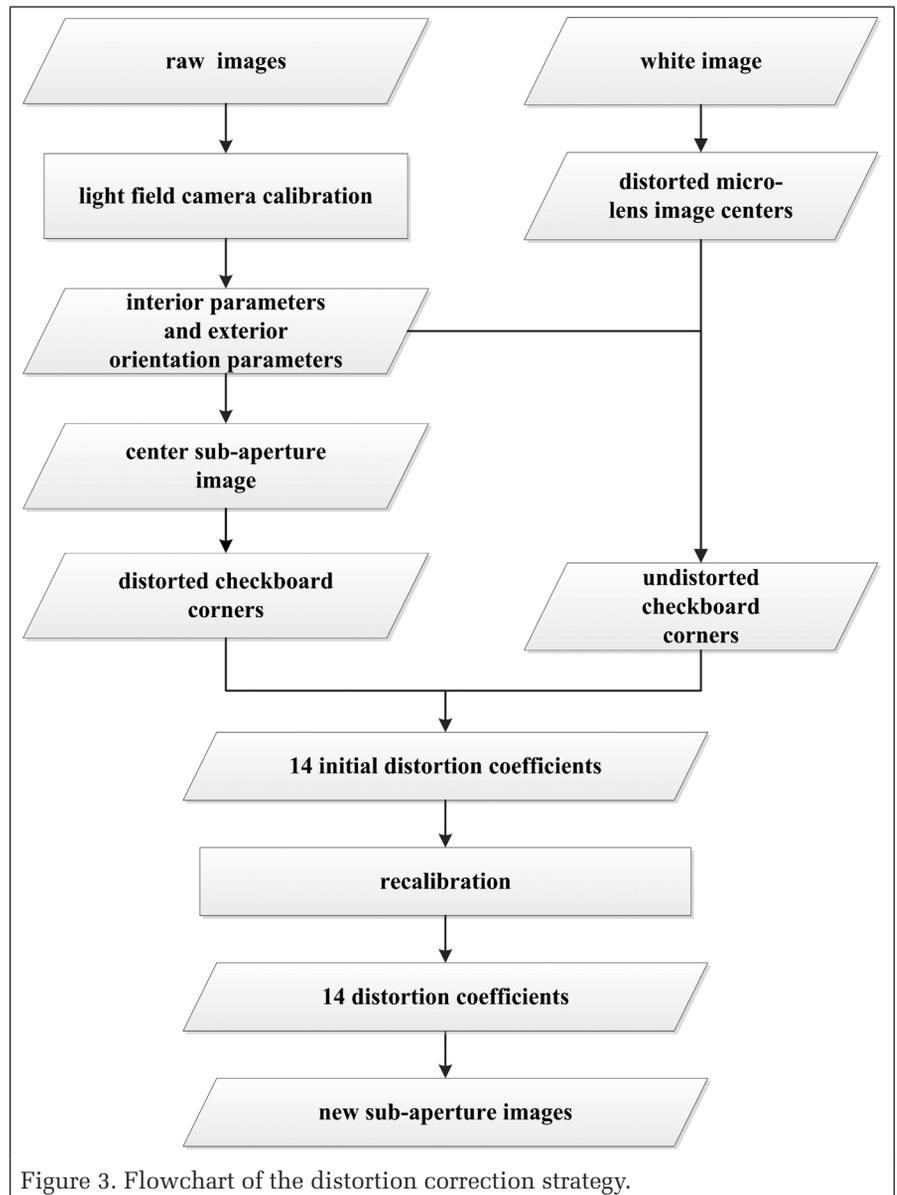
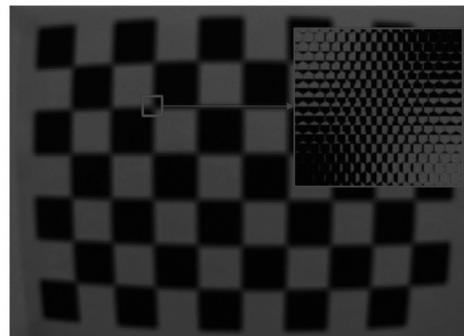


Figure 3. Flowchart of the distortion correction strategy.



(a)



(b)

Figure 4. (a) The Lytro Illum camera, and (b) An example of a raw image from Bok's dataset for calibration

corresponds to a pixel (the red grid in Figure 5b) in a sub-aperture image. Taking the same pixel underneath each virtual micro-lens, sub-aperture images from different angles can be obtained. The corresponding relationship between the coordinate (u_s, v_s) in the sub-aperture image and the virtual micro-lens image center (u_c, v_c) in the raw image is as follows:

$$\begin{aligned} u_c &= (u_s - 1) * s + u_0 \\ v_c &= (v_s - 1) * s + v_0 \end{aligned} \quad (12)$$

where s is a scaling factor whose value equals the space between the centers of two virtual micro-lenses (here, $s = 14$). Moreover, (u_0, v_0) is the coordinate of the first center of the virtual micro-lenses in the CCD array (here, $u_0 = 7.5 \text{ pixel}$, $v_0 = 7.5 \text{ pixel}$).

The value of a pixel underneath virtual micro-lens can be calculated by barycentric interpolation among the pixels with same displacements from adjacent micro-lens image centers underneath three nearest micro-lenses to the virtual micro-lens. For example, as shown in Figure 5a, the value of the green grid underneath the virtual micro-lens (red dashed box) can be calculated by barycentric interpolation among the green pixels underneath three nearest micro-lenses to the virtual micro-lens. In practice, the size of the sub-aperture images, which corresponds to the number of the virtual micro-lenses, can be arbitrary. However, we set the size of the sub-aperture images to be the same as the size of the raw image scaled by the diameter of a real micro-lens (14 pixels).

Acquisition of Undistorted and Distorted Checkboard Corners
The white image shown in Figure 6 is an image taken through a white diffuser, or of a white scene by the light field camera

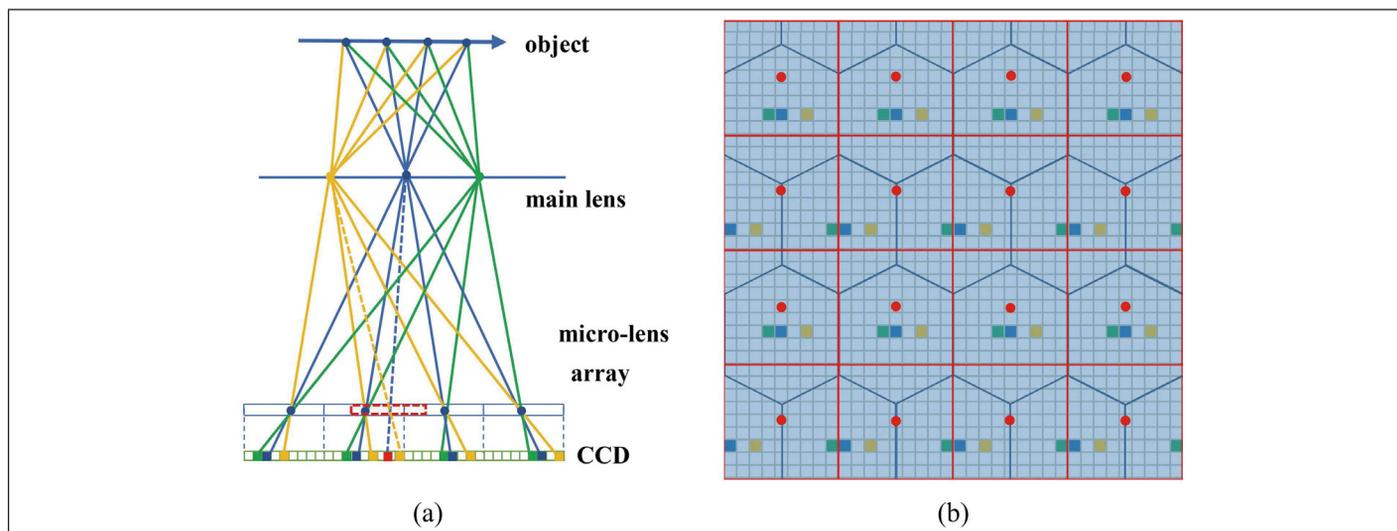


Figure 5. (a) Imaging model of light field camera (side view). The blue rectangles represent micro-lenses. The red dashed box represents a virtual micro lens; and (b) Overhead view of (a). The aqua grid represents the CCD pixels, and the overlaid blue hexagons represent the micro-lenses. The red dots are our virtual micro-lens image centers. The green, blue, and yellow squares represent the pixels with same displacements from adjacent micro-lens image centers, respectively. Each red grid cell represents one pixel in the sub-aperture image.

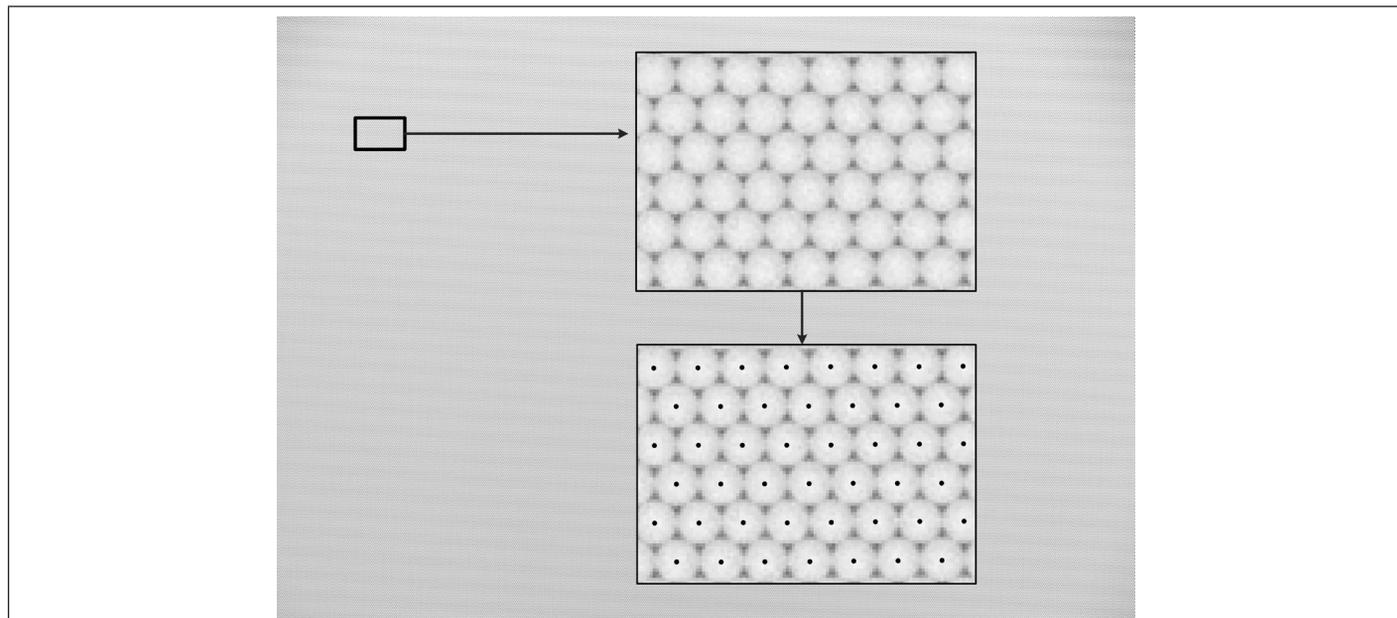


Figure 6. A white image showing the effect of vignetting. The brightest spots in the micro-lens image are extracted as the micro-lens image centers.

(Dansereau *et al.*, 2013). Because of vignetting, the brightest spot in each micro-lens image approximates its center.

The distorted micro-lens images centers can be extracted using the white image (Bok *et al.*, 2014). The undistorted micro-lens images centers are calculated by removing distortion from the distorted checkerboard corners with distortion coefficients k_1 and k_2 . The distorted micro-lens images centers and checkerboard corners on the raw image are shown in Figure 7

The undistorted checkerboard corners (\hat{u}_i, \hat{v}_i) in the raw image can be calculated using the equation:

$$\begin{bmatrix} \hat{u}_i \\ \hat{v}_i \end{bmatrix} = \begin{bmatrix} \hat{u}_{cl} + \Delta u_i \\ \hat{v}_{cl} + \Delta v_i \end{bmatrix} \quad (13)$$

where $(\hat{u}_{cl}, \hat{v}_{cl})$ are the corresponding undistorted micro-lens images centers, and $(\Delta u_i, \Delta v_i)$ are the corresponding displacements of the checkerboard corners from the undistorted micro-lens centers. In order to get $(\hat{u}_{cl}, \hat{v}_{cl})$, the displacements Δu and Δv varied from -4 pixels to 4 pixels in intervals of 0.5 pixels. Then, the corresponding undistorted micro-lens image centers (\hat{u}_c, \hat{v}_c) are computed according to the following equation, which can be got from Equation 1:

$$\begin{bmatrix} \hat{u}_c \\ \hat{v}_c \end{bmatrix} = \begin{bmatrix} \frac{f_x X_c - \Delta u (K_1 Z_c + K_2)}{Z_c} \\ \frac{f_y Y_c - \Delta v (K_1 Z_c + K_2)}{Z_c} \end{bmatrix} \quad (14)$$

where (X_c, Y_c, Z_c) are the 3D coordinates of checkerboard corners in the camera coordinate system, which can be calculated by transforming the world coordinate (X_w, Y_w, Z_w) of the checkerboard corners with the rotation matrix \mathbf{R} and the translation vector \mathbf{t} from the above calibration:

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = \mathbf{R} \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} + \mathbf{t} \quad (15)$$

The calculated (\hat{u}_c, \hat{v}_c) may not be included in the undistorted micro-lens image centers. Therefore, the undistorted micro-lens image centers nearest to (\hat{u}_c, \hat{v}_c) are chosen as $(\hat{u}_{cl}, \hat{v}_{cl})$. The corresponding $(\Delta u_i, \Delta v_i)$ are calculated by Equation 1. Finally, (\hat{u}_i, \hat{v}_i) are obtained by Equation 13).

The distorted checkerboard corners on the raw image (\hat{u}_a, \hat{v}_a) are extracted by matching the center sub-aperture image with a raw image based on least-squares matching. The checkerboard corners to be matched in the center sub-aperture image are extracted through corner detection. The initial checkerboard corners to be matched in the raw image are derived by adding distortion on the undistorted checkerboard corners (\hat{u}_i, \hat{v}_i) with distortion parameters k_1 and k_2 .

All extracted distorted checkerboard corners (\hat{u}_a, \hat{v}_a) are shown in Figure 8.

Recalibration with the 14-parameter Distortion Model

The light field camera records the information of the rays from multiple directions, and therefore various types of aberrations need to be considered to correct distortions. A geometric model that explicitly describes how an aberrated

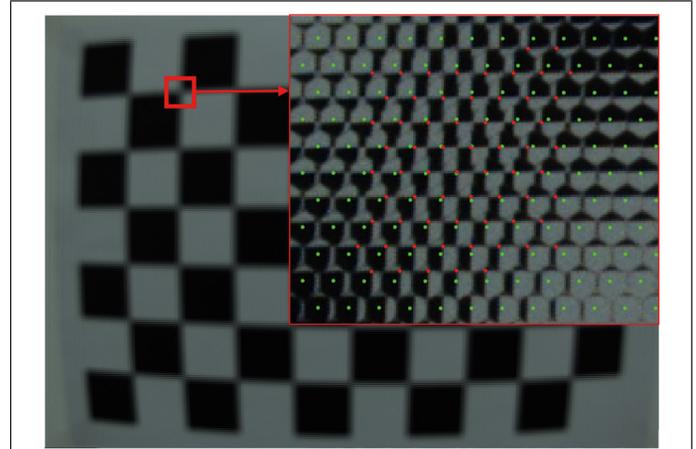


Figure 7. A raw image of a checkerboard pattern. The green dots represent the distorted micro-lens image centers. The red dots represent the distorted checkerboard corners.

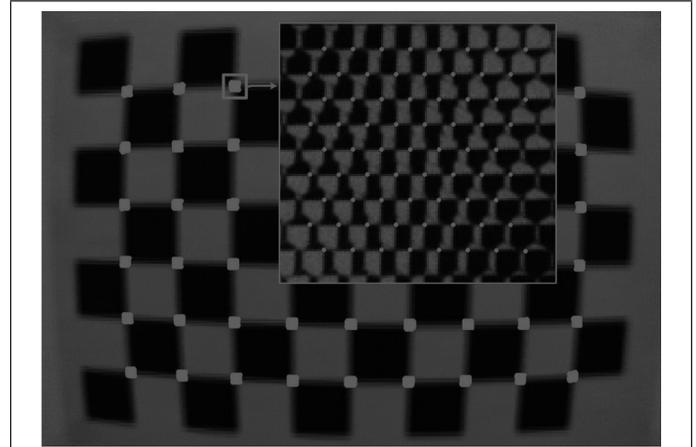


Figure 8. Distorted checkerboard corners in the raw image.

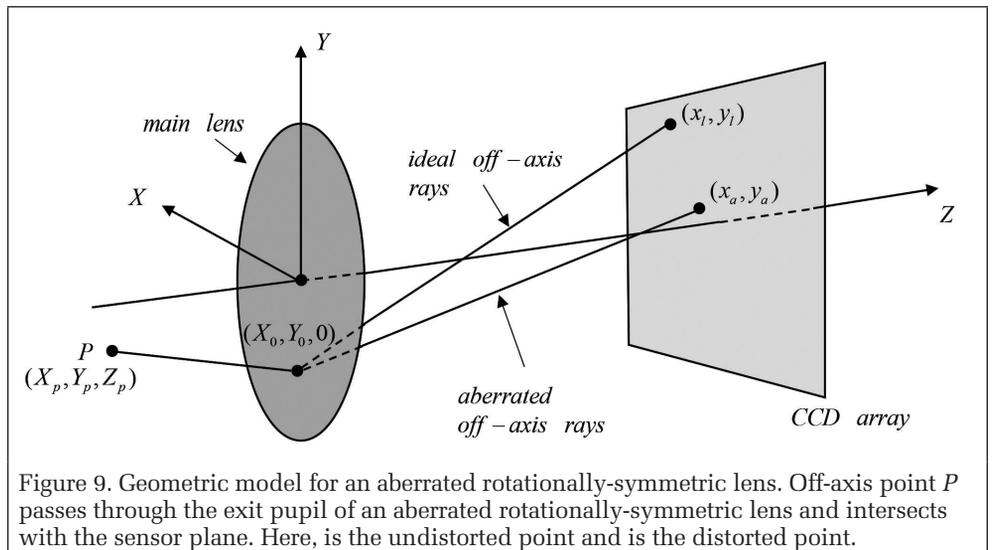


Figure 9. Geometric model for an aberrated rotationally-symmetric lens. Off-axis point P passes through the exit pupil of an aberrated rotationally-symmetric lens and intersects with the sensor plane. Here, (x_i, y_i) is the undistorted point and (x_a, y_a) is the distorted point.

Table 1. Primary aberrations (Mahajan, 1984).

Wave aberration terms	Aberration name
$W_1(e, l) = e l^4$	Spherical
$W_2(e, l) = e l^2 (e \cdot l)$	Coma
$W_3(e, l) = e \cdot l ^2$	Astigmatism
$W_4(e, l) = e l^2 l ^2$	Field curvature
$W_5(e, l) = l ^2 (e \cdot l)$	Distortion

Table 2. Secondary aberrations (Mahajan, 1984).

Wave aberration terms	Aberration name
$W_6(e, l) = e l^6$	Spherical*
$W_7(e, l) = e l^4 (e \cdot l)$	Coma*
$W_8(e, l) = e l^2 (e \cdot l)^2$	Astigmatism*
$W_9(e, l) = (e \cdot l)^3$	Arrows
$W_{10}(e, l) = e l^4 l ^2$	Lateral spherical
$W_{11}(e, l) = e l^2 l ^2 (e \cdot l)$	Lateral coma
$W_{11}(e, l) = e l^2 l ^2 (e \cdot l)$	Lateral astigmatism
$W_{13}(e, l) = e l^2 l ^4$	Lateral field curvature*
$W_{14}(e, l) = l ^4 (e \cdot l)$	Lateral distortion*

*The word "secondary" is associated with these aberrations, e.g., secondary spherical.

Table 3 Correlation coefficients for the aberration parameters.

Parameter 1	Parameter 2	Correlation coefficient (%)
α_1	α_4	62.93
α_1	α_6	89.19
α_2	α_5	66.16
α_2	α_7	73.70
α_2	α_{14}	57.10
α_3	α_4	51.47
α_3	α_8	45.27
α_3	α_{12}	83.80
α_4	α_6	50.35
α_4	α_{12}	42.57
α_4	α_{13}	81.73
α_5	α_{14}	94.24
α_6	α_{10}	46.23
α_8	α_{10}	55.10
α_9	α_{11}	56.52
α_{11}	α_{14}	42.76
α_{12}	α_{13}	57.45

rotationally-symmetric lens bends a ray is introduced (Figure 9). There are two basic rays: an ideal off-axis ray that describes how a thin lens bends a ray at a general point $(X_0, Y_0, 0)$ on the exit pupil, and the aberrated off-axis ray that passes through the same point on the pupil but deviates from the ideal one.

For a rotationally-symmetric lens, its aberrations include primary, secondary, and tertiary aberrations (Mahajan, 1984). It was found experimentally that the first two aberrations are sufficient to correct the light field image accurately. Aberration $V(e, l)$ can be modeled as a function of the pupil point $e = (X_0, Y_0)$ and the undistorted point $l = (x_l, y_l)$ (Mahajan, 1984). The function can be written as a power series (Tang and Kutulakos, 2013), our 14-parameter distortion model is represented as:

$$V(e, l) = \sum_k a_k \frac{\partial W_k(e, l)}{\partial e} \quad (16)$$

where $a_k (k = 1, 2, \dots, 14)$ are the distortion coefficients, $W_k(e, l)$ are the wave aberration terms.

For primary (or Seidel) aberration, the aberration terms are listed in Table 1.

For secondary (or Schwarzschild) aberrations, the aberration terms are listed in Table 2.

The aberration items $\frac{\partial W_k(e, l)}{\partial e}$ are obtained by taking

the partial derivative of the wave aberration items $W_k(e, l)$ with respect to e . For instance, the term $\frac{\partial W_5(e, l)}{\partial e} = \left(\frac{\partial W_5(e, l)}{\partial X_0}, \frac{\partial W_5(e, l)}{\partial Y_0} \right)$,

where $\frac{\partial W_5(e, l)}{\partial X_0} = x_l(x_l^2 + y_l^2)$, and $\frac{\partial W_5(e, l)}{\partial Y_0} = y_l(x_l^2 + y_l^2)$. The

term $\frac{\partial W_{14}(e, l)}{\partial e} = \left(\frac{\partial W_{14}(e, l)}{\partial X_0}, \frac{\partial W_{14}(e, l)}{\partial Y_0} \right)$, where

$\frac{\partial W_{14}(e, l)}{\partial X_0} = x_l(x_l^2 + y_l^2)^2$, and $\frac{\partial W_{14}(e, l)}{\partial Y_0} = y_l(x_l^2 + y_l^2)^2$.

The coefficients a_5 and a_{14} corresponds to the traditional radial distortion coefficients k_1 and k_2 .

The high correlations between parameters may affect the numerical solution and the estimated parameters (Clarke and Fryer, 1998). We analyze the correlation between the relevant parameters. Correlation coefficients are obtained directly from the covariance matrix using a given set of observation equations and the functional model (Equation 16). The corners used to compute the distortion coefficients using Equation 16 are 37,233 in total from 20 images captured from different distances and orientations. The parameters which showed a high level of correlation are shown in Table 3.

From Table 3, the radial distortion parameters a_5 and a_{14} have the highest correlation coefficients 94.24%. There are different extents of correlation between other parameters. But the high correlation coefficients don't imply that one parameter can dropped without consequence (Clarke *et al.*, 1998). It is found experimentally that RMSE values of projection errors can achieve minimization using all of 14 parameters to correct distortion. Therefore, the 14-parameter distortion model is applied in the paper.

With the 14-parameter distortion model, the field light camera is recalibrated using distorted checkerboard corners, 14 initial distortion coefficients, interior parameters $(K_1, K_2, f_x, f_y, c_x, c_y)$ and exterior orientation parameters.

The initial values of 14 distortion parameters are obtained through minimizing the following objective function:

$$G = \sum_{\min} (x_a - x_l - V_x(e, l)^2 + y_a - y_l - V_y(e, l)^2) \quad (17)$$

where $V(e, l)$ is calculated according to Equation 16. For a lenslet light field camera, the point $e = (X_0, Y_0)$ on the exit pupil can be calculated from the following equation (Bok *et al.*, 2014):

$$\begin{bmatrix} x_a \\ y_a \end{bmatrix} = \begin{bmatrix} \dot{u}_a / f_x \\ \dot{v}_a / f_y \end{bmatrix}, \quad (18)$$

where $(\Delta u, \Delta v)$ are the displacements from the micro-lens image centers. The distorted checkerboard corners $l_a = (x_a, y_a)$ and undistorted checkerboard corners $l = (x_l, y_l)$ are computed using:

$$\begin{bmatrix} x_a \\ y_a \end{bmatrix} = \begin{bmatrix} \dot{u}_a / f_x \\ \dot{v}_a / f_y \end{bmatrix}, \begin{bmatrix} x_l \\ y_l \end{bmatrix} = \begin{bmatrix} \dot{u}_l / f_x \\ \dot{v}_l / f_y \end{bmatrix}, \quad (19)$$

The extracted distorted checkerboard corners in raw image are denoted by $m = [\dot{u}_a, \dot{v}_a]^T$. The coordinates of the checkerboard corners in the world coordinate system is denoted by $m = [X_w, Y_w, Z_w]^T$. If the images used to calibrate are taken from N poses and there are C checkerboard corners in each image used to calibrate, the distortion coefficients $(a_1, a_2, \dots, a_{14})$ can be obtained through an iterative optimization by minimizing the following objective function:

$$\sum_{i=1}^N \sum_{j=1}^C m_{ij} - \check{m}(K_1, K_2, f_x, f_y, c_x, c_y, a_1, a_2, \dots, a_{14}, R_i, t_i, M_j)^2 \quad (20)$$

where \check{m} is the projection of checkerboard corner M_j in image i calculated by adding distortion $V(e, l)$ on the undistorted checkerboard corners (\dot{u}_l, \dot{v}_l) , and (\dot{u}_l, \dot{v}_l) are calculated by Equation 13, $(\dot{u}_{cl}, \dot{v}_{cl})$ in Equation 13 is calculated by removing distortion from the corresponding distorted micro-lens image centers with radial distortion parameters a_5 and a_{14} . The corresponding $(\Delta u_l, \Delta v_l)$ in Equation 13 are calculated by Equation 1. For Equation 20, it is a nonlinear minimization problem, which is solved with the Levenberg-Marquardt algorithm as implemented in More (1978).

Improved Depth Estimation

It is found experimentally that there are a number of outliers on depth estimation in the weak texture region due to the influence of narrow baseline, noise and aliasing for a lenslet light field camera. The proposed depth estimation method is developed based on Jeon's method (Jeon *et al.*, 2015) with a

simple improvement strategy to reduce the influence of outliers on depth estimation. The cost volume in Jeon's algorithm is refined. In our proposed algorithm, the center sub-aperture image is divided into strong and weak texture regions by a strong texture confidence measure. Costs are only calculated in strong texture regions. With the cost volume, the multi-label optimization propagates the correct depth in the strong texture regions to the weak texture regions. At last, the disparity refinement (Yang *et al.*, 2007) is used to recover a non-discrete depth map.

Strong Texture Confidence

The strong texture confidence measure C_e is used to test which regions of the center sub-aperture images the depth estimation seems promising. We define it as:

$$C_e(u, v) = \sum_{u' \in \mathcal{N}(u, v)} |I_c(u, v) - I_c(u', v)| + \sum_{v' \in \mathcal{N}(u, v)} |I_c(u, v) - I_c(u, v')| \quad (21)$$

where $\mathcal{N}(u, v)$ is a 1D window in the center sub-aperture image I_c . The size of the neighborhood can be small (seven pixels in our experiment) as it is supposed to measure only the local color variation.

In order to get more reliable strong texture region, the guided filter (He *et al.*, 2013) is applied to smooth C_e with the center sub-aperture image. Then, a threshold is set to C_e , resulting in a binary confidence mask M_e , which marks the strong texture regions as shown in Figure 10.

Cost Volume Computation

Our algorithm only computes the matching costs for the pixels in the center sub-aperture images with $M_e(u, v) = 1$. The matching costs for the pixels with $M_e(u, v) = 0$ is set to a constant value. The cost volume $C(u, v, l)$ computation is defined as:

$$C(u, v, l) = \alpha C_A(u, v, l) + (1 - \alpha) C_G \quad (22)$$

where $\alpha \in (0, 1)$ is the balancing parameter. The sum of absolute differences $C_A(u, v, l)$ is defined as:

$$C_A(u, v, l) = \begin{cases} \sum_{(u, v) \in \Omega} \min(|I_c(u, v) - I_o(u + d_s, v + d_t)|, \tau_1) & \text{if } M_e(u, v) = 1 \\ \tau_1 & \text{if } M_e(u, v) = 0 \end{cases} \quad (23)$$

where I_c is the center sub-aperture image, I_o are the other sub-aperture images, and τ_1 in the first equation in Equation 23 represents a truncation used for removing the bad matches, $C_A(u, v, l)$ in the second one in Equation 23 one can be arbitrary constant value. Here, it is set to τ_1 . Moreover, Ω is a small square region centered at (u, v) . The sub-pixel shift vectors d_s, d_t of the multi-view sub-aperture images are defined as:

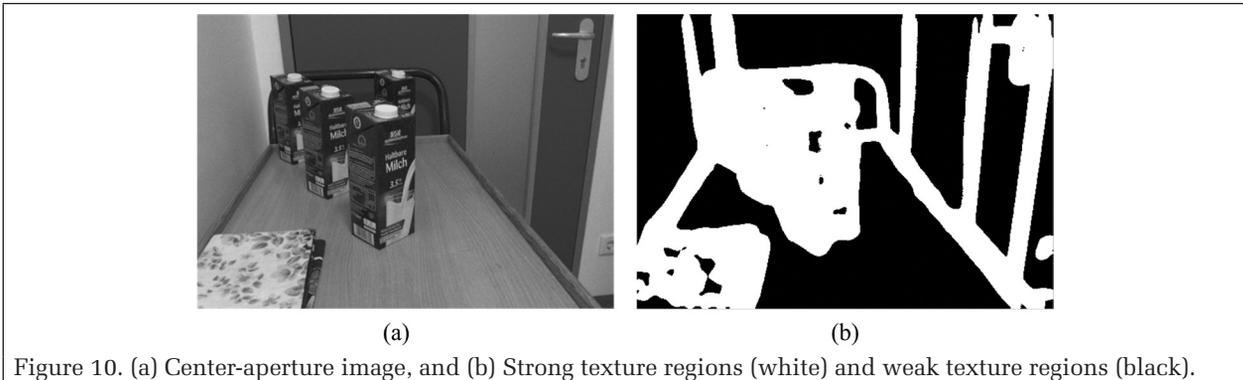


Figure 10. (a) Center-aperture image, and (b) Strong texture regions (white) and weak texture regions (black).

$$\begin{aligned} d_s &= kl(s - s_c) \\ d_t &= kl(t - t_c) \end{aligned} \quad (24)$$

where k is the unit of label in pixels, (s_c, t_c) is the center view, and (s, t) are the other views. The sum of gradient differences C_G is defined as

$$C_G(u, v, l) = \begin{cases} \sum_{(u, v) \in \Omega} \beta \min(|\nabla_u I_c(u, v) - \nabla_u I_o(u + d_s, v + d_t)|, \tau_2) \\ + (1 - \beta) \min(|\nabla_v I_c(u, v) - \nabla_v I_o(u + d_s, v + d_t)|, \tau_2) & \text{if } M_e(u, v) = 1 \\ \tau_2 & \text{if } M_e(u, v) = 0 \end{cases} \quad (25)$$

where $\beta \in (0, 1)$ control the relative importance of two direction gradient differences, and τ_2 in the first equation in Equation 25 represents a truncation used for removing the bad matches, $C_G(u, v, l)$ in the second one in Equation 25 can be arbitrary constant value. Here, it is set to τ_2 . In order to alleviate the coarsely scattered unreliable matches, the guided filter (He *et al.*, 2013) is used in every cost slice to get the refined cost volume $C'(u, v, l)$.

Multi-label Optimization

Multi-label optimization using graph cuts (Kolmogorov *et al.*, 2002) is performed to propagate and correct the disparities. The optimal disparity map is obtained through minimizing

$$l_r = \operatorname{argmin} \sum_{(u, v)} C'(u, v, l(u, v)) + \lambda_1 \sum_{(u, v) \in I} \|l(u, v) - l_a(u, v)\| + \lambda_2 \sum_{(u', v') \in N_c} \|l(u, v) - l(u', v')\| \quad (26)$$

where l is the center sub-aperture image, $C'(u, v, l(u, v))$ is the refined matching cost using guided filter, $(\|l(u, v) - l_a(u, v)\|)$ is the data fidelity, and $\|l(u, v) - l(u', v')\|$ is the local smoothness, which is used to propagate the correct disparities to the weak texture regions.

Disparity Refinement

After the multi-label optimization, the discrete disparity map is obtained. In order to get a continuous disparity map, the

Table 4. Interior parameters with Bok's method for Bok's dataset.

Parameters	K_1	$K_2(mm)$	$f_x(pixel)$	$f_y(pixel)$	$c_x(pixel)$	$c_y(pixel)$	k_1	k_2
Value	3.373	2428.955	7134.867	7128.613	3842.742	2719.563	0.334	0.105

Table 5. Distortion coefficients with our method for Bok's dataset.

Coefficient	Value	Coefficient	Value
α_1	7.4569e-6	α_8	-5.6999e-6
α_2	2.9391e-5	α_9	8.2009e-6
α_3	-3.2600e-5	α_{10}	-2.1190e-6
α_4	5.3997e-5	α_{11}	-7.5619e-6
α_5	-0.3228	α_{12}	1.0090e-4
α_6	-1.3753e-6	α_{13}	3.9430e-5
α_7	-5.6920e-6	α_{14}	0.1730

Table 6. Comparisons of reprojection errors based on Bok's (left) and our (right) models using Bok's dataset (unit: pixel).

Δv	Δu				
	-3	-2	0	2	3
-3	0.255 0.151	0.207 0.154	0.177 0.164	0.195 0.158	0.236 0.142
-2	0.218 0.154	0.185 0.157	0.167 0.154	0.174 0.168	0.196 0.159
0	0.199 0.161	0.185 0.156	0.168 0.137	0.175 0.166	0.182 0.172
2	0.230 0.159	0.203 0.164	0.184 0.161	0.187 0.163	0.216 0.155
3	0.278 0.163	0.236 0.163	0.203 0.163	0.224 0.158	0.270 0.154

method proposed by Yang *et al.* (2007) is applied. The continuous disparity l^m is calculated as:

$$l^m = l - \frac{C(l_+) - C(l_-)}{2(C(l_+) + C(l_-) - 2C(l))} \quad (27)$$

where $l_+ = l + 1$ and $l_- = l - 1$ are the adjacent cost slices of l , l is the discrete disparity with the minimal cost, and l^m is the disparity with the minimum cost, which is derived from the least square quadratic fitting over three costs: $C(l_-)$, $C(l)$ and $C(l_+)$.

Experimental Results

Distortion Correction Result

After calibrating the camera using raw images, the sub-aperture images can be generated based on the interior parameters and the distortion parameters. The exterior orientation parameters from the calibration result transform the checkerboard coordinate to the camera coordinate. The transformation from the camera coordinate to micro-lens image centers coordinate is computed using Equation 14. The transformation from micro-lens image centers coordinate to sub-aperture coordinate is computed using the inverse transformation of Equation 12). Corner features are extracted from sub-aperture images independently, and the root-mean-squared error (RMSE) values of projection errors can be computed. To verify the effectiveness of our distortion model, we applied our correction method and Bok's method (Bok *et al.*, 2014) on the Lytro Illum calibration dataset provided by Bok (2017). The calibration result obtained by Bok's method and our method for Bok's datasets are shown in Tables 4 and 5. Table 6 shows the comparisons of reprojection errors based on Bok *et al.* (2014) and our models using Bok's dataset.

In addition, we apply our correction method and Bok's method (Bok *et al.*, 2014) on our dataset to further verify the effectiveness of our distortion model. Our dataset is captured using a Lytro Illum camera at distances from 210 mm to 250 mm, which includes 12 images of a checkerboard pattern whose grid size is 7.22 mm. The camera is calibrated using Bok's geometric calibration toolbox (Bok, 2017), and the ray re-projection error in 3D is 0.088 mm. The calibration result is shown in Table 7. The distortion coefficients obtained by our method for our datasets are shown in Tables 8 and 9 and show the comparisons of reprojection errors based on Bok *et al.* (2014) and our models using our dataset.

From Table 6 with Table 9, the projection errors of the sub-aperture images corrected by our method are clearly smaller. Especially, the reprojection errors of the sub-aperture images with larger displacement are significantly improved using our correction method. The RMSE in Table 6 for Bok's and ours is 0.208 and 0.155 pixel, respectively. The RMSE in Table 9 for Bok's and ours is 0.635 and 0.424 pixel, respectively. The average projection error for the two datasets has decreased by about 30% with our correction method.

Depth Estimation Result

The effectiveness of our distortion correction method and depth estimation

method is qualitatively and quantitatively evaluated on the Hazirbas' dataset (Hazirbas et al., 2017), which is a real-scene indoor benchmark composed of 4D light-field images captured by Lytro Illum camera and with ground truth depth obtained from a registered RGB-D sensor. In order to verify the effectiveness of our distortion model, the sub-aperture images corrected by Bok's (Bok et al. 2014) and our methods are used to estimate depth with the state-of-the-art algorithm developed by Jeon et al. (2015). Jeon's source code is released on the website (https://sites.google.com/site/hgjeoncv/home/depthfromlf_cvpr15). Moreover, in order to verify the effectiveness of our depth estimation method, we have implemented it to compare with the method developed by Jeon et al. (2015) by using the same sub-aperture images corrected by our correction method. We fix the user-defined parameters for a fair comparison for the different correction method and dedicate to obtain the best results by optimizing parameters for different depth estimation method. The code is running in MATLAB on an Intel i7 3.4 GHZ and 24 GB memory PC. The parameters experimentally perform well with $\alpha = 0.5$, $\tau_1 = 0.5$, $\tau_2 = 0.5$, label_number=40. The parameters λ_1 and λ_2 are empirically selected as $\lambda_1 = 2$, $\lambda_2 = 0.009$ for Bok's and our correction method, $\lambda_1 = 4$, $\lambda_2 = 0.01$ for our depth estimation method. The absolute error map $|d_{estimate} - d_{groundtrue}|$ and the corresponding error distribution histogram (the error of greater than 500 mm is set to 500 mm) are shown in Figure 11. Bok's plus Jeon's represents Bok's correction method and Jeon's depth estimation method. Ours plus Jeon's represents our correction method and Jeon's depth estimation method. Ours plus depth represents our correction method and our depth estimation method.

In order to compare the results in detail, parts of the images in the red and green rectangles shown in the error maps are magnified in Figure 11. Comparing the red squares in Figure 11b and 11c in the three scenes, it can be seen that the depth errors in the strong texture areas decrease using our correction methods, which reflects that our correction method is better than Bok's in correcting the distortions. Comparing to the green squares in Figure 11b and 11c using depth estimation method of Jeon *et al.*, we can find that using our method, the depth errors shown in Figure 11d decrease also in the weak texture regions. This is because our depth estimation method reduces the effects of outliers in the weak texture regions. Accordingly, the reduction in depth error due to our distortion correction method and the error reduction caused by our depth estimation method are both reflected in the error distribution histogram. As can be seen from the histograms shown in Figure 11, the concentrations trend of the bins toward smaller error in the error distribution histograms is obvious by using our methods. From histograms in Figure 11b

Table 7. Interior parameters derived by the calibration method (Bok et al., 2014).

Parameters	K_1	$K_2(mm)$	$f_x(pixel)$	$f_y(pixel)$	$c_x(pixel)$	$c_y(pixel)$	k_1	k_2
Value	-147.167	26708.272	23148.233	23116.376	3879.220	2710.382	0.686	-0.717

Table 8. Interior parameters with our method for our dataset.

Coefficient	Value	Coefficient	Value
a_1	-1.5343e-5	a_8	1.6164e-6
a_2	1.3736e-5	a_9	8.7593e-6
a_3	2.1034e-4	a_{10}	-1.9066e-5
a_4	7.3973e-4	a_{11}	3.1570e-5
a_5	-0.3484	a_{12}	0.0030
a_6	-2.7023e-6	a_{13}	-0.0013
a_7	-1.2063e-6	a_{14}	1.7863

Table 9. Comparisons of reprojection errors based on Bok's (left) and our (right) models using our dataset (unit: pixel).

Δv	Δu				
	-3	-2	0	2	3
-3	0.686 0.443	0.585 0.484	0.595 0.396	0.663 0.493	0.734 0.521
-2	0.691 0.460	0.607 0.389	0.511 0.337	0.593 0.379	0.717 0.522
0	0.686 0.370	0.605 0.422	0.277 0.274	0.519 0.408	0.649 0.432
2	0.662 0.364	0.618 0.348	0.572 0.401	0.567 0.343	0.708 0.547
3	0.791 0.477	0.685 0.383	0.551 0.397	0.632 0.521	0.781 0.373

Table 10. RMSE of the absolute error (unit: mm).

	Bok's+Jeon's	Ours+Jeon's	Ours+depth
Scene 1	232.38	200.27	82.89
Scene 2	220.97	181.24	128.40
Scene 3	122.91	118.04	95.72

Table 11. Median of the absolute error (unit: mm).

	Bok's+Jeon's	Ours+Jeon's	Ours+depth
Scene 1	203.72	147.50	62.66
Scene 2	171.50	88.30	81.06
Scene 3	101.11	84.46	73.12

to 11c, c, and from 11c to 11d, the changes of the bins concentrations verify the effectiveness of error reduction of both our distortion correction and depth estimation methods.

The quantitative evaluations of our distortion correction and depth estimation methods are performed by computing RMSE and median of the absolute error. The RMSE for the above three scenes are shown in Table 10. The median of the absolute errors for the above three scenes are shown in Table 11.

From the middle two columns of Table 10, we can find that the magnitude of RMSE reduction is not prominent, the average RMSE reduction of three scenes is about 12%; this is because the distortion correction only reduces the depth error in the strong texture region. The last two columns of Table 10 show that with the same distortion correction method, our depth estimation method improves the accuracy of depth estimation in weak texture regions, resulting in the obvious reduction of depth errors (about 42% for the average RMSE reduction of three scenes). The images we used in Hazirbas' datasets are captured at the range of 80~150 cm. The RMSE for our depth estimation method act in accordance with the theoretical analysis in Figure 2a, which verifies the correctness of our method. From Table 11, we can find the depth errors move toward smaller values using our correction and depth estimation methods, which is consistent with movement trend in the errors distribution histograms shown in Figure 11. All the experiment results demonstrate the effectiveness of our

proposed distortion correction and depth estimation methods.

Though our methods produce obvious improvements on depth estimation. There are still some small regions with large depth errors in our depth estimation results. For example, from Figure 11d in scene 1, we can find that the depth errors within the region shown in the blue square are large. This is because that the region is occluded all around and the correct depth cannot be propagated into it. From Figure 11d in scene 2, the large error within the region shown in the blue square are mainly resulted from the influence of the shadow. The effects caused by occlusion and shadow are our future research work.

Conclusions

In this paper, we proposed a distortion correction and depth estimation method for lenslet light field cameras. The improvements include two aspects: (i) a novel 14-parameter distortion model was used to correct the light field camera images and was demonstrated to be superior to the traditional two parameter radial distortion model; and (ii) a depth estimation method is proposed to reduce the impact of outliers in weak texture regions. In addition, a new algorithm was proposed for sub-aperture image generation and the theoretical measurement error equation for the lenslet light field camera is derived. Distortion correction and depth estimation experiments were carried out to validate the feasibility and effectiveness of our proposed method. The experimental results show that our method outperforms the state-of-art depth estimation method with respect to accuracy.

According to the theoretical measurement analysis and experimental results, the unfocused plenoptic camera is suitable for measurement within a couple of meters at a millimeter level precision. In the future, with the foreseeable development of the light field camera, e.g., by increasing the number of micro-lenses and/or increasing the size of CCDs, lens improvement, we believe that it will be widely used in more applications.

Acknowledgments

This research is funded by National Natural Science Foundation of China (Project no. 41471388), National Basic Research Program of China (Project no. 2013CB733202), and National Key Research and Development Program of China (Project no.2016YFB0502102). The authors thank Yunsu Bok and

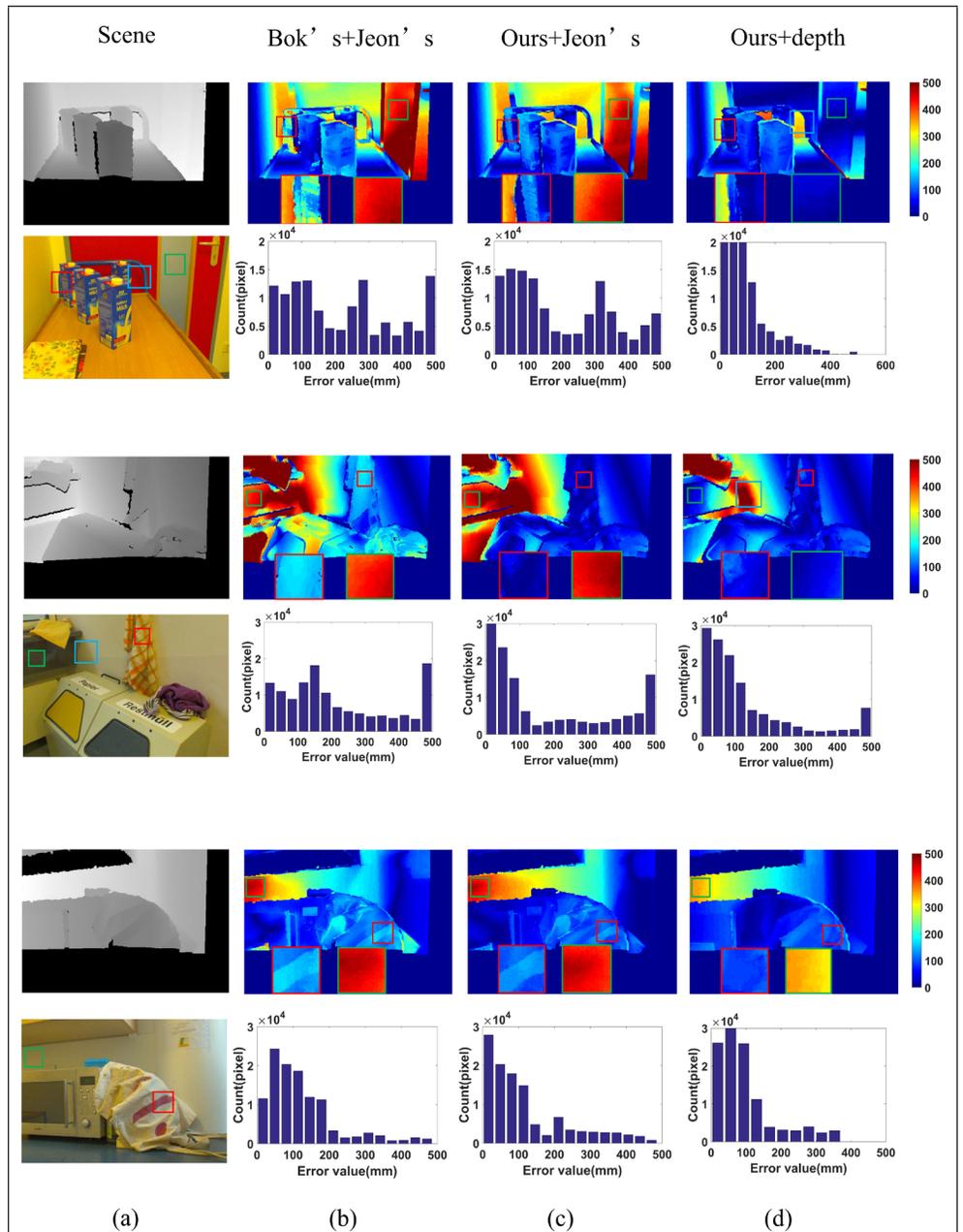


Figure 11. Absolute error map (unit:mm) and error distribution histograms. Every two rows represent a separate scene. For each two rows: (a) center sub-aperture image (bottom) and ground truth depth map (top); (b-c) Absolute error maps (top) and error distribution histograms (bottom) from Bok's and our correction methods both using Jeon's depth estimation method; and (d) Absolute error map (top) and error distribution histogram (bottom) using our correction method and our depth estimation method.

Hae-Gon Jeon of the Korea Advanced Institute of Science and Technology (KAIST) for help with their projection model and for making the calibration toolbox available on the Web.

References

- Ackermann, F., 1984. Digital image correlation: Performance and potential application in photogrammetry, *The Photogrammetric Record*, 11(64):429-439.
- Adelson, E.H, and J.Y.A. Wang, 1992. Single lens stereo with a plenoptic camera, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):99-106.

- Apelt, F., D. Breuer, Z. Nikoloski, M. Stitt, and F. Kragler, 2015. Phytotyping4D: a light-field imaging system for non-invasive and accurate monitoring of spatio-temporal plant growth, *The Plant Journal*, 82(4): 693-706.
- Bok, Y., Yunsu Bok's Homepage, URL: https://sites.google.com/site/yunsubok/lf_geo_calib (last data accessed: 03 October 2018).
- Bok, Y., H.G. Jeon, and I.S. Kweon, 2014. Geometric calibration of micro-lens-based light-field cameras using line features, *Proceedings of the European Conference on Computer Vision*, 05-12 September, Zurich, Switzerland, pp. 47–61.
- Brown, D.C., 1966. Decentering Distortion of Lenses, *Photogrammetric Engineering*, V. XXXII, Number 3, pp. 444-462.
- Chen, C., H. Lin, Z. Yu, S.B. Kang, and J. Yu, 2014. Light field stereo matching using bilateral statistics of surface cameras, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 23-28 June, Columbus, Ohio, pp. 1518-1525.
- Cho, D., M. Lee, S. Kim, and Y.W. Tai, 2013. Modeling the calibration pipeline of the lytro camera for high quality light-field image reconstruction, *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 01-08 December, Sydney, NSW, Australia, pp. 3280–3287.
- Clarke, T.A., and J.G. Fryer, 1998. The development of camera calibration methods and models, *The Photogrammetric Record*, 16 (91):51-66.
- Clarke, T.A., X. Wang, and J.G. Fryer, 1998. The principal point and CCD cameras, *The Photogrammetric Record*, 16(92):293-312.
- Dansereau, D.G., I. Mahon, O. Pizarro, S.B. Williams, 2011. Plenoptic flow: Closed-form visual odometry for light field cameras, *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 25-30 September, San Francisco, California, pp.4455-4462.
- Dansereau, D.G., O. Pizarro, and S.B. Williams, 2013. Decoding, calibration and rectification for lenselet-based plenoptic cameras, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 23-28 June, Portland, Oregon, pp. 1027–1034.
- Di, K., and R. Li, 2007. Topographic mapping capability analysis of Mars exploration rover 2003 mission imagery, *Proceedings of the 5th International Symposium on Mobile Mapping Technology (MMT 2007)*, Padua, Italy, May 28-31. (CD-ROM).
- Dong, F., S.H. Ieng, X. Savatier, R. Etienne-Cummings, and R. Benosman, 2013. Plenoptic cameras in real-time robotics, *The International Journal of Robotics Research*, 32(2):206-217.
- Fryer, J.G., 1996. Camera Calibration (K.B. Atkinson, editor), *Close Range Photogrammetry and Machine Vision*, Bristol, Whittle Publishing, pp. 156-179.
- Hazirbas, C., L. Leal-Taixé, and D. Cremers, 2017. Deep depth from focus, *arxiv preprint arXiv: 1704. 01085*.
- He, K., J. Sun, and X. Tang, 2013. Guided image filtering, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(6):1397-1409.
- Heinze, C., S. Spyropoulos, S. Hussmann, and C. Perwass, 2016. Automated robust metric calibration algorithm for multifocus plenoptic cameras, *IEEE Transactions on Instrumentation and Measurement*, 65(5):1197-1205.
- Honauer, K., O. Johannsen, D. Kondermann, and B. Goldluecke, 2015. A dataset and evaluation methodology for depth estimation on 4D light fields, *Proceedings of the Asian Conference on Computer Vision*, 21-23, Nov 2015, Taipei, Taiwan, pp. 19-34.
- Jeon, H.G., J. Park, G. Choe, J. Park, Y. Bok, Y.W. Tai, and I.S. Kweon, 2015. Accurate depth map estimation from a lenslet light field camera, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 07-12 June 2015, Boston, Massachusetts, pp. 1547–1555.
- Kim, C., H. Zimmer, Y. Pritch, A. Sorkine-Hornung, and M. Gross, 2013. Scene reconstruction from high spatio-angular resolution light fields, *Proceedings of the ACM Transactions on Graphics (ACM SIGGRAPH)*, 32(4),73-1.
- Kolmogorov, V., and H. Zabih, 2002. Multi-camera scene reconstruction via graph cuts, *Proceedings of the European Conference on Computer Vision (ECCV)*, 28-31 May, Copenhagen, Denmark, pp. 8-40.
- Liang, C., and K., Ramamoorthi, 2015. A light transport framework for lenslet light field cameras, *ACM Transactions on Graphics (TOG)*, 34(2):16.
- Lin, H., C. Chen, S.B. Kang, and J. Yu, 2015. Depth recovery from light field using focal stack symmetry, *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 07-13 December 2015, Santiago, Chile, pp. 3451–3459.
- Lumsdaine, A., and T. Georgiev, 2009. The focused plenoptic camera, *Proceedings of the IEEE International Conference on Computational Photography (ICCP)*, 16-17 April, San Francisco, California, pp. 1-8.
- Lytro camera, URL: <http://www.lytro.com> (last date accessed: 03 October 2018).
- Mahajan, V.N. 1984. *Optical Imaging and Aberrations: Ray Geometrical Optics*, SPIE press, Bellingham, Washington, 484 p.
- More, J., 1978. The Levenberg-Marquardt algorithm: Implementation and theory, *Numerical Analysis*, Springer, Berlin, Heidelberg, pp.105-116.
- Ng, R., M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, 2005. Light field photography with a hand-held plenoptic camera, *Computer Science Technical Report (CSTR)*, 2(11):1-11.
- Perwaß, C., and L. Wietzke, 2012. Single lens 3d-camera with extended depth-of-field, *Proceedings of Human Vision and Electronic Imaging XVII*, Burlingame, California.
- Raytrix 3d light field camera technology, URL: <http://www.raytrix.de/>. (last date accessed: 03 October 2018).
- Sabater, N., M. Seifi, V. Drazic, et al., 2014. Accurate disparity estimation for plenoptic images. *Proceedings of the European Conference on Computer Vision*, Springer International Publishing, 05-12 September, Zurich, Switzerland, pp. 548-560.
- Strobl, K., M. Lingenauber, 2016. Stepwise calibration of focused plenoptic cameras, *Computer Vision and Image Understanding*, 145:140-147.
- Tao, M.W., S. Hadap, J. Malik, and R. Ramamoorthi, 2013. Depth from combining defocus and correspondence using light field cameras, *Proceedings of International Conference on Computer Vision (ICCV)*, 01-08 December 2013, Sydney, NSW, Australia, pp. 673-680.
- Tosic, I., and K. Berkner, 2014. Light field scale-depth space transform for dense depth estimation, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 23-28 June, Columbus, Ohio, pp. 435-442.
- Tang, H., and K.N. Kutulakos, 2013. What does an aberrated photo tell us about the lens and the scene?, *Proceedings of International Conference on Computational Photography (ICCP)*, 19-21 April. Cambridge, Massachusetts, pp. 1-10.
- Wanner, S., and B. Goldluecke, 2013. Variational light field analysis for disparity estimation and super-resolution, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(3):606-619.
- Yang, Q., R. Yang, J. Davis, and D. Nistér, 2007. Spatial-depth super resolution for range images. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 17-22 June, Minneapolis Minnesota, pp.1-8.
- Yu, Z., X. Guo, H. Ling, A. Lumsdaine, and J. Yu, 2013. Line assisted light field triangulation and stereo matching, *Proceedings of International Conference on Computer Vision (ICCV)*, 1-8 December, Sydney, NSW, Australia, pp. 2792-2799.
- Zeller, N., C.A. Noury, F. Quint, C. Teulière, U. Stilla, and M. Dhome, 2016. Metric calibration of a focused plenoptic camera based on a 3D calibration target, *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Science*, Volume III-3, XXIII ISPRS Congress, Prague, Czech Republic, pp. 449-456.
- Zhang, Z., Y. Liu, and Q. Dai, 2015. Light field from micro-baseline image pair, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 07-12 June, Boston, Massachusetts, pp. 3800-3809.

Analysis and Correction of Digital Elevation Models for Plain Areas

Cristian Guevara Ochoa, Luis Vives, Erik Zimmermann, Ignacio Masson, Luisa Fajardo, and Carlos Scioli

Abstract

Water movement modeling in plain areas requires digital elevation models (DEMs) adequately representing the morphological and geomorphological land patterns including the presence of civil structures that could affect water flow patterns. This has a direct effect on water accumulation and water flow direction. The objectives of this work were to analyze, compare and improve DEMs so surface water movement in plain areas could be predicted. In order to do that, we evaluated the accuracy of a digital elevation data set consisting in 4064 points measured with a differential global positioning system (GPS) in a plain area of central Buenos Aires province. Three DEMs were analyzed: (1) the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), (2) the Shuttle Radar Topography Mission (SRTM) and (3) the Advanced Land Observing Satellite with the Phased Array Type L-Band Synthetic Aperture Radar (ALOS PALSAR). Several topographic attributes (i.e., height, surface area, land slope, delimitation of geomorphological units, civil structures, basin boundaries and streams network) and different interpolation methods were analyzed. The results showed that both the SRTM and the ALOS PALSAR DEMs had a ± 4.4 m root mean square error (RMSE) in contrast to the ASTER DEM which had a ± 9 m RMSE. Our analysis proved that the best DEM representing the study area is the SRTM. The most suitable interpolation methods applied to the SRTM were the inverse distance weighting and the ANUDEM, whereas the spline method displayed the lowest vertical accuracy. With the proposed method we obtained a DEM for the study area with a ± 3.2 m RMSE, a 33% error reduction compared to the raw DEM.

Introduction

Earth's surface plays a fundamental role in the modeling of hydrological processes (Wilson and Galán 2000; Wilson 2012;

Cristian Guevara Ochoa, Luis Vives, Ignacio Masson, and Luisa Fajardo are with the Instituto de Hidrología de Llanuras "Dr. Eduardo Jorge Usunoff", IHLLA, República de Italia 780 C.C. 47 (B7300) Azul, Buenos Aires, Argentina (cguevara@ihlla.org.ar).

Cristian Guevara Ochoa and Erik Zimmermann are with the Concejo Nacional de Investigaciones Científicas de Argentina, CONICET, Av. Rivadavia 1917, (C1033AAJ) Ciudad Autónoma de Buenos Aires, Argentina.

Erik Zimmermann is with the Centro Universitario Rosario de Investigaciones Hidroambientales, CURIHAM, Riobamba 245 bis (2000) Rosario, Santa Fe, Argentina.

Ignacio Masson and Luisa Fajardo are with the Comisión de Investigaciones Científicas de la provincia de Buenos Aires, CIC, Calle 526 e/10 y 11, La Plata, Buenos Aires, Argentina.

Carlos Scioli is with the Centro de estudios de variabilidad y cambio climático, CEVARCAM, Facultad de Ingeniería y Ciencias Hídricas, Universidad Nacional del Litoral, Paraje el pozo C.C. 217, Ruta Nacional N° 168, Km 472, Santa Fe, Argentina.

Eric *et al.* 2014). Satellite technologies are currently being developed for capturing topographic information through digital elevation models (DEMs). Through the use of DEMs, geomorphological (Hutchinson *et al.* 2001) and hydrological properties (Jarihani *et al.* 2015) can be analyzed. These include soil moisture (Ludwig and Schneider 2006; Gao *et al.* 2016), flood impact (Sanders 2007; Tarekegn *et al.* 2010; Gichamo *et al.* 2012; Yan *et al.* 2014; Wurl *et al.* 2014), soil stability, potential erosion, precipitation retention, channel shape, land depressions, etc. (Barnes *et al.* 2014).

DEMs have been used for a variety of environmental applications, such as modeling water processes. The most commonly used data for DEMs are radar and radiometric reflection data such as Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) data (Ica and Hook 2002; Pachri *et al.* 2013; Carrascal *et al.* 2013). There are many examples of DEMs using Shuttle Radar Topography Mission (SRTM) data (e.g., Gesch *et al.* 2006; Lin *et al.* 2013; Sharma and Tiwari 2014). Likewise, other applications use Advanced Land Observing Satellite with the Phased Array Type L-Band Synthetic Aperture Radar (ALOS PALSAR) data (e.g., Pontes *et al.* 2017; Hidayat *et al.* 2017).

Currently, the models that quantify surface water movements require a better resolution. Therefore, semi-distributed and distributed models appeared to reproduce various processes that occur in the water balance, such as evaporation, runoff and infiltration. According to Easton *et al.* (2008), Vaze *et al.* (2010), and Guevara (2015), the relief represents an important aspect for which the detailed topography adequately representing the hydrology is needed. This is because when hydrological studies are carried out in plain areas DEMs do not properly represent the hydrological characteristics of the surface (Callow *et al.* 2007).

According to Chaubey *et al.* (2005), the DEM resolution conditions model calibration, having an effect on basin delineation and total surface area (Martz and Jong 1988), drainage network prediction (Chen *et al.* 2012), subcatchment classification and slope.

The relationship between topography and water flow is less clear in plain areas, where low height gradients and depressions make surface flow tracing more complex (Gallant and Dowling 2003). This difficulty is aggravated by the civil structures that contribute to the topographic uncertainty in these areas, since any structure with a height even below 1 m has a significant impact on the surface flow in terms of its direction and quantity (Guevara 2015). On the other hand, the water flow in plain areas does not strictly follow the topography, especially in water excess conditions.

In plain areas, DEMs do not adequately represent the channel geometry. This leads to misinterpretations of the hydraulic factor of water transport, estimated from the

Photogrammetric Engineering & Remote Sensing
Vol. 85, No. 3, March 2019, pp. 209–219.
0099-1112/18/209–219

© 2019 American Society for Photogrammetry
and Remote Sensing
doi: 10.14358/PERS.85.3.209

cross-sectional geometry of the channel, which negatively affects water levels, flow rates, time, flood wave velocity and simulated flood dynamics (Tarekegn *et al.* 2010).

Therefore, most of the time it is necessary to correct the DEMs since they generally do not represent the natural and artificial channels, main and secondary roads, and other structures that could be a few meters wide and deep (Scioli 2009; Guevara 2015). These errors could be corrected by adjusting the height of the DEM using multiple sources of information, such as topographic maps, satellite images, interpolation methods and vectors taken by differential global positioning system (GPS) on roads, depressions and channels.

DEMs provide essential spatial data for hydrological and hydrodynamic modeling. For this reason, the resolution of the DEM and the method with which elevation data were interpolated affect the quality of the results.

There is a wide variety of interpolation methods either deterministic (e.g., inverse distance weighting, natural neighbor, spline and ANUDEM) or geostatistical (e.g., kriging). However, there are few published studies on the effectiveness of interpolators applied to the same data set. Most studies were carried out by Reuter *et al.* (2007) and Yang and Hodler (2013). There are different techniques for DEM correction in hydrological modeling. Among them the ANUDEM interpolation stands out as a response to the elimination of data noise (Callow *et al.* 2007). This method interpolates surface data through points, lines and polygons (Hutchinson 1989).

There is no agreement on which are the best methods of interpolation, if deterministic methods or geostatistical methods. Several authors such as Paredes *et al.* (2013) and Yang and Hodler (2013) recommended that DEM validations should be carried out locally to make better decisions as to which interpolation method works best. Some studies establish that geostatistical methods are more reliable for the prediction of height values (Guo *et al.* 2010; Arun 2013), while others, such as Reuter *et al.* (2007), come to a different conclusion suggesting that deterministic methods are better for predicting height values.

The objectives of this study were: (a) To evaluate, by means of using different statistical procedures and graphics, the vertical accuracy of ASTER, SRTM and ALOS PALSAR DEMs in plain areas, accounting for the morphological, geomorphological and civil structures in the terrain, in order to get the best DEM representing topographic variability (b) to compare vertical accuracy using different interpolation methods applied to the best DEM and (c) to establish guidelines for the correction of DEMs in plain areas.

Study Area

The study area was located in the center of the province of Buenos Aires (36.8°–37.3° S, 58.8°–60.1° W). With a surface area of 2,725 km² (Figure 1), it covers Del Azul upper creek basin. The main stream origins in the town of Chillar (60 km south of the city of Azul) and its main tributaries are the Videla stream with a 120 km² sub-basin and the Santa Catalina stream with a 138 km² sub-basin.

The geomorphology of the study area is dominated by positive reliefs in the hills zone (Tandilia system) and decreases towards the north, where there is a depressed zone, with very soft and low reliefs and <1% slopes. This area lacks of a developed drainage system. In periods of water excess this causes the entire surface of the landscape to become flooded due to the low hydraulic capacity of the channels (Guevara 2015).

The study area is characterized by being shaped by the action of wind, due to the low morphometric potential and the fine granulometry of the soils, which makes it prone to wind erosion. The concentrated action of wind deflation in these zones is capable of excavating closed depressions, known as deflation hollows, which play a very important role in the water flow and storage in these areas.

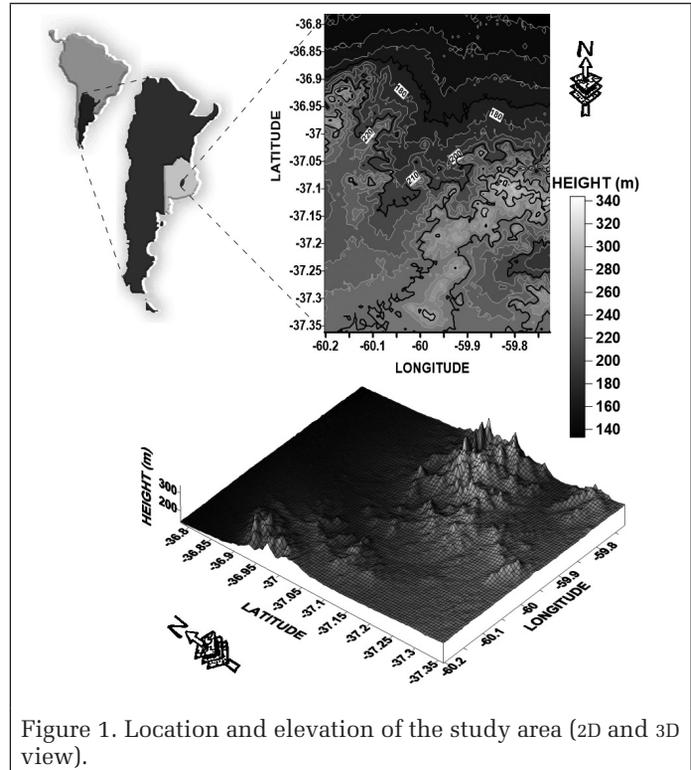


Figure 1. Location and elevation of the study area (2D and 3D view).

The water moves northeast from the Tandilia hills system and it is mainly lost by evaporation (Zabala *et al.* 2015). This is one of the regions of the world with the lowest morphometric potential and for this reason surface runoff has a shallow but extensive coverage. In this landscape, deflation hollows play a very important role in the movement of water, since these are connected forming a runoff pathway parallel to the flow of the main channel.

Methodology

In this section, we present the observed data set we worked with, we describe the used DEMs, we explain how we calculated the flow and accumulation of water for the drainage network, we describe the deterministic and geostatistical interpolation methods we used, and finally, we explain how the accuracy of the DEM was determined.

Data Collection Using a GPS

To evaluate the vertical accuracy of the DEMs, a differential GPS unit was used. This type of GPS measures height using the EGM96 projection (Lemoine *et al.* 1998). This is a datum that takes into account the height of the geoid and through which the vertical accuracy of a DEM could be evaluated.

A Thales Promark 3 differential GPS unit was the particular equipment used for the measurements. This equipment allows the post-processing of the measured points using Global Navigation Satellite System (GNSS) software to correct for vertical errors and bring them down to the cm range. To analyze the vertical accuracy of the DEMs in the study area, eight control points 10 km apart from each other were used and a topographic survey accounting for 4064 points was carried out.

Differential GPS data points were taken on main and secondary roads, land depressions and in the stream channel (bathymetry). Figure 2 shows the location of the GPS measured points and the control points, and some graphic examples of how points were distributed in order to account for the particular topographic features of the study area.

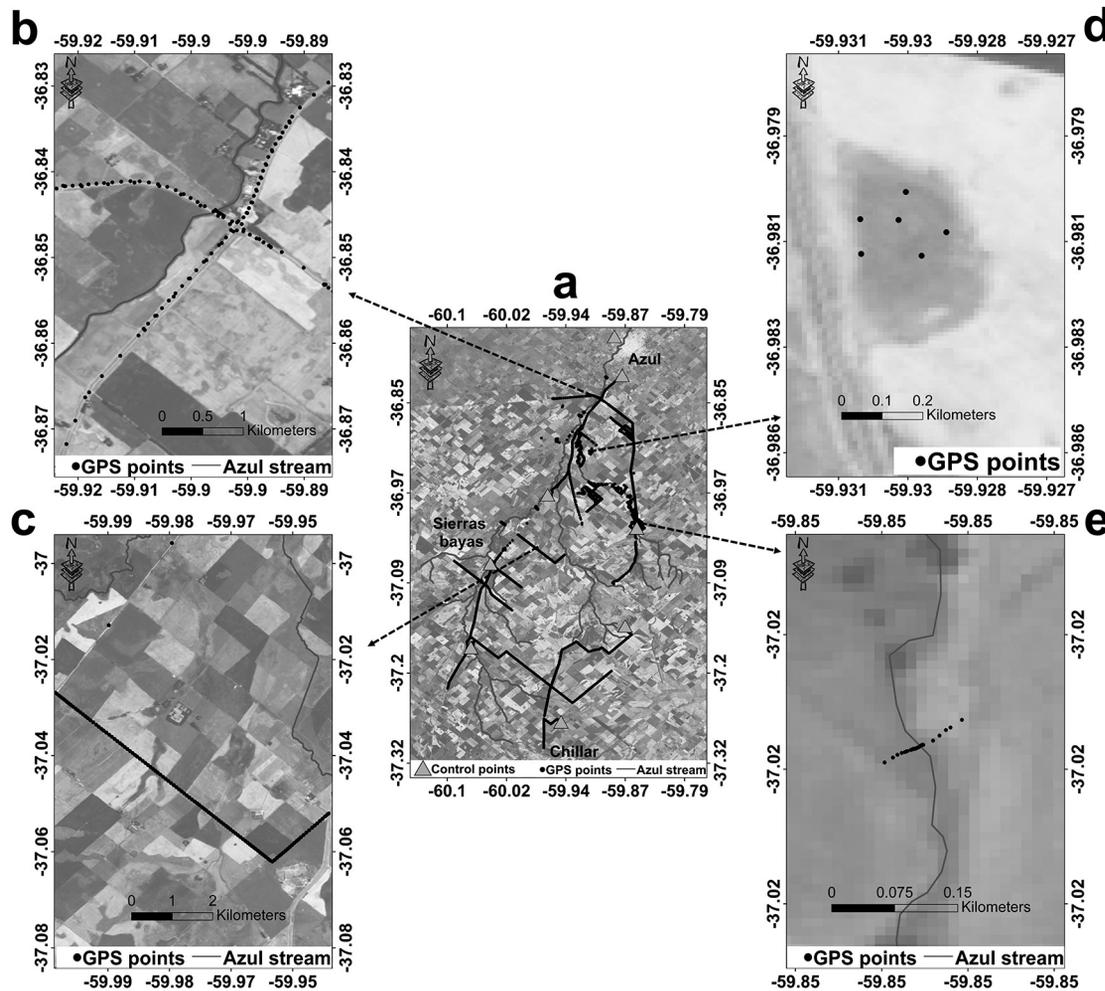


Figure 2. (a) Location of the points measured using differential GPS and the control points for the study area, (b) GPS points measured on main roads, (c) GPS points measured on secondary roads, (d) GPS points measured in a deflation hollow, (e) GPS points measured in the stream channel (bathymetry). Source: satellite image SPOT 5.

Description of the Used DEMs

Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER)

This DEM created by the Ministry of Economy, Trade and Industry of Japan (METI), uses an infrared spectral band generating stereo images, with a base-height ratio of 0.6 (Tachikawa *et al.* 2011) that covers the world land surface from 83° N to 83° S (Abrams *et al.* 2015). It has a spatial reference of 1 arc second (~30 m) worldwide coverage. The digital elevation model (GDEM) Version 2 was released in October 2011 (three years after its predecessor, Version 1), along with the addition of 200 000 new images for the stereoscopic process, several anomalies were corrected, and its overall accuracy was improved. This DEM has a ± 17 m vertical accuracy (Mukherjee *et al.* 2013; Rexer and Hirt 2014; Robinson *et al.* 2014).

Shuttle Radar Topography Mission (SRTM)

This DEM is a cooperative project between the United States National Aeronautics and Space Administration (NASA) and the United States Department of Defense's National Imagery and Mapping Agency (NIMA). The mission was designed to use a radar interferometer to produce a digital elevation model of the Earth's surface between approximately 60° N and 56° S, that is about 80% of the Earth's land mass (Rabus *et al.* 2003). This mission was carried out for 11 days in February

2000. This DEM has a spatial reference of 1 second of arc (~30 m) and has a ± 16 m vertical precision (Sun *et al.* 2003; Kellndorfer *et al.* 2004; Gesch 2006).

Advanced Land Observing Satellite (ALOS PALSAR)

The DEM of the Advanced Land Observing Satellite (ALOS) was created by the Japan Space Exploration Agency (JAXA) and uses the synthetic L-band type aperture sensor (PALSAR) which operates at a band width of 14 to 28 MHz. This DEM has a spatial resolution of 12.5 m and covers an area of 156.25 m² per pixel (Kimura and Ito 2000; Rosenqvist *et al.* 2007; Chu and Lindenschmidt 2017). It has a WGS84 horizontal reference datum and the vertical datum is modified by means of the gravitational model of the earth (EGM08) that takes into account the vertical reference of the geoid (Pavlis *et al.* 2012).

Calculation of Drainage Network and Delimitation of the Basin

These tasks were performed using the Archydro extension of ArcGIS (ESRI 2011) for the geographical region corresponding to the Del Azul upper creek basin. The Archydro extension uses the D8 algorithm (Tarboton 1997) with which the water flow in each pixel is evaluated considering its eight neighboring pixels. Once flow directions are calculated, the flow field (i.e., the area contributing to each cell or pixel of the grid) is

quantified to define the resulting direction of the flow and the accumulation of water. This allows the delimitation of the basin boundaries in order to compare the different contributing areas discretized by each DEM.

Interpolation Methods

An element that introduces error into DEMs is the interpolation method used to generate the model. Interpolation methods are procedures used to predict values at locations lacking sampling points. These methods are based on the principle of spatial autocorrelation or spatial dependence (Childs 2004). Previous studies have presented results that do not agree with each other as to the accuracy of the different interpolation methods used for the generation of the DEMs (Wechsler and Kroll 2006; Reuter *et al.* 2007; Yang and Hodler 2013; Arun 2013). In order to contribute to this discussion and to improve our understanding of the error introduced by the interpolation methods in DEMs, our study evaluates and compares the precision of DEMs generated using two different interpolation methods: the deterministic and the geostatistical methods.

Deterministic Methods

Spatial interpolation is a procedure used to calculate the value of a variable in a certain spatial position, knowing the values of that variable in other positions of the space.

Inverse Distance Weighting (IDW): According to (Bartier and Keller 1996), this method uses a higher allocation or gives a higher weight to the nearest point with this weight decreasing as the distance increases, depending on the power coefficient.

$$z_{x,y} = \frac{\sum_{i=1}^n Z_i W_i}{\sum_{i=1}^n W_i} \quad (1)$$

where:

$z_{x,y}$: It is the estimate of the height at the point (x,y) .
 z_i : It represents the control value for the i th sampling point.
 w_i : It is a weight that determines the relative importance of the individual control point Z_i in the interpolation procedure.
 n : It is the number of points.

The value of the missing data is estimated based on a weighted average of the measured data and a weight is assigned at each point depending on its location. The equation is:

$$w_i = d_{x,y,i}^{-\beta} \quad (2)$$

where:

$d_{x,y,i}$: It is the distance between $z_{x,y}$ and z_i .
 β : It is a user-defined exponent, in this case it is 2 because the inverse of the squared distance was used.

Natural neighbor: Sibson (1981) developed a weighted average function, called natural neighbor interpolation. This method is based on the Voronoi diagram, applied to a discrete set of spatial points. Natural neighbor interpolation finds a subset of samples closest to an input query point and applies weights to them based on proportional areas to interpolate a value. It is a spatial autocorrelation method, which decreases the standard deviation (STD) of values (Wechsler and Kroll 2006).

This type of interpolation is similar to IDW, it calculates the point value depending on the position of the neighboring point.

$$Z_{x,y} = \sum_{i=1}^n w_i Z(s_i) \quad (3)$$

where:

$Z_{x,y}$: It is the estimate of the height at the point (x,y) .
 w_i : It is a weight that determines the relative importance of point Z_i .
 $Z(s_i)$: It is the measured value at the location $i(x_i, y_i, z_i)$.
 n : It is the number of points.

Spline: This method calculates the height value at a point, with a mathematical function that minimizes curvature of the surface, resulting in a smoothed surface, which passes exactly through the points in the sample (Childs 2004).

This method uses low-grade polynomials, thus avoiding undesirable fluctuations in most of the applications found, by interpolating using high-grade polynomials (Wahba and Wendelberger 1980). In our study, we used the regularized spline method that incorporates the first derivative of the slope.

The spline interpolation is calculated with the following formula:

$$Z_{(x,y)} = T_{(x,y)} + \sum_{j=1}^n \lambda_j R(r_j) \quad (4)$$

where:

$Z_{(x,y)}$: It is the estimate of the height at a point (x,y) .
 i : It is the point number.
 n : It is the total number of points.
 λ_j : It is a coefficient determined by solving a system of linear equations.
 r_j : It is the distance from point (x,y) to point i .
 $T_{(x,y)}$ and (r) : These are defined differently according to the selected option.

The regularized spline interpolation option was the method used in this study. The calculation used $T_{(x,y)}$ and (r) from the following equations:

$$T(x,y) = a1 + a2x + a3y \quad (5)$$

where:

a_j : The coefficients are determined by solving the system of linear equations.

And (r) is determined by the following formula:

$$R(r) = \frac{1}{2\pi} \left\{ \frac{r^2}{4} \left[\ln \left(\frac{r}{2\tau} \right) + c - 1 \right] + \tau^2 \left[K_0 \left(\frac{r}{\tau} \right) + c + \ln \left(\frac{r}{2\pi} \right) \right] \right\} \quad (6)$$

where:

r : It is the distance between the point and the sample.
 τ : It is the weight of the parameter.
 K_0 : It is the modified Bessel function.
 c : It is a constant equal to 0.577215.

ANUDEM: This algorithm was developed by Hutchinson (1989). It constitutes a morphological approach to the interpolation of digital elevation models and uses a specially designed interpolation technique to create a surface that represents more accurately the drainage system.

It is an adaptive mesh that is commonly used for the calculation of digital elevation models in a regular grid. It includes the implementation of the algorithm and the application that imposes a drainage structure connected to an interpolated DEM (Hutchinson *et al.* 2011).

This coupling is performed with an iterative interpolation technique based on finite differences, which optimizes computational efficiency through Childs (2004) minimization.

$$\sum_{i=1}^n \left[\frac{z_i - f(x_i, y_i)}{w_i} \right]^2 + \Lambda J(f) \quad (7)$$

where:

$J(f)$: It is the roughness of the terrain as a function of the first and second derivative of f .

Λ : It is the positive parameter that softens said roughness.

f : It is an unknown bivariate function representing a finite difference mesh.

Parameter Λ is chosen so that the sum of squares of the residuals in the equation above is equal to n , which can only be achieved by an iterative interpolation method for which, the slope of each cell is available. That is, the ANUDEM method generates a low-resolution DEM and interpolates improving the resolution until reaching the solution.

$$z_i = f(x_i, y_i) + \varepsilon_i (i=1, \dots, n) \quad (8)$$

where:

n : It is the number of elevation samples.

ε_i : It is the random error with mean 0 and standard deviation given by:

$$w_i = h s_i \sqrt{12} \quad (9)$$

where:

h : It is the cell size of the mesh.

S_i : It is the slope of the cell associated with the sample.

Geostatistical Method (Kriging)

This method developed by Matheron (1965) is a local spatial interpolation, based on variogram theory and structural analysis (Zhang *et al.* 2015). It is a spatial correlation of the data points measured by the variogram function. The peculiarity of this type of interpolation is that it has the ability to produce a prediction surface but also provides some degree of certainty or accuracy of predictions.

The kriging interpolation method uses the following formula:

$$Z_{(x,y)} = \sum_{i=1}^n \lambda_i Z(s_i) \quad (10)$$

where:

$Z_{(x,y)}$: It is the estimate of the height at a point (x,y) .

$Z(s_i)$: It is the measured value at location $i(x_i, y_i, z_i)$.

λ_i : It is an unknown weight n for the measured value at position i .

n : It is the number of measured values.

The method used in our study is ordinary kriging. It is the most general and most widely used kriging method. It assumes the mean is constant and unknown. The variogram model used is spherical, which shows a progressive decrease in the autocorrelation (and in the increase of semivariance) of the distance.

Evaluation of Vertical Accuracy of DEMs

To evaluate the accuracy of the different digital elevation models, Wechsler and Kroll (2006) and Shortridge and Mesina (2011) recommended three different types of statistics

(root mean square error, coefficient of determination and standard deviation).

According to Wilson and Galán (2000) the most used statistic for the evaluation of digital elevation model's accuracy is the root mean square error (RMSE). The RMSE represents the standard deviation of the differences between the calculated values and the observed values in the sample. These individual differences are known as residuals, when calculations are carried out on the sample data used for the estimation of prediction errors. The RMSE is a suitable measure of accuracy, but only to compare prediction errors for a particular variable and not among variables, since it is scale dependent. The desired RMSE value is 0, indicating that the method did not produce errors.

Another method to evaluate is the use of the coefficient of determination (R^2). It is a statistic that describes the proportion of variance in the observed data. It is the square of the Pearson correlation coefficient, which varies between 0 and 1. When R^2 adopts high values it indicates a smaller variance of the error, in general, values above 0.7 are considered acceptable. The R^2 has been widely used for hydrological evaluation, although it is more sensitive to extreme values.

The standard deviation is a measure of dispersion also used. It is the square root of the variance of the variable, and the deviation is the distribution that presents the data around the arithmetic mean.

Results

The detailed results of the study are presented in the following subsections. We first defined the general differences between the three models at representing elevation, area and slope for the study area. Then we analyzed in a detailed scale the differences of the three models at reflecting the topographic attributes of the terrain such as geoforms, civil structures and stream channel shape (bathymetry). We also examined the direction of the water flow and water accumulation in order to characterize the drainage network and to delineate basin boundaries. Thereafter, the DEM that best represented the topography of our study area was chosen. Finally, in order to improve the accuracy of the chosen model, we examined different interpolation methods.

Differences Among DEMs

Figure 3 shows the height differences among the three DEMs. For the ASTER model, height ranged from 371 to 115 m. For the SRTM model height ranged from 370 to 114 m. And for the ALOS PALSAR model height ranged from 368 to 114 m. When we evaluated the vertical differences among ASTER, SRTM and ALOS PALSAR models for the plain zone, the linear regression analysis we performed revealed a strong significant correlation between the GPS measured height and the height provided by each of the three DEMs. However, this correlation was stronger both for the SRTM and the ALOS PALSAR models ($R^2 = 0.98$ for both) than for the ASTER model ($R^2 = 0.93$) (Figure 4). It is interesting to note that in all three cases the slope value of the regression line was close to 1.

Figure 5 shows the boxplot for the GPS and the DEMs data. In this graph we can differentiate the frequency distribution, the symmetry of the data in terms of elevation and its atypical values. The SRTM and ALOS PALSAR DEMs showed a height distribution with a symmetry and a degree of homogeneity comparable with that of the GPS data. Unlike these two DEMs, the ASTER model showed a more asymmetric data distribution when compared to the GPS data distribution as well as more dispersion and atypically higher values.

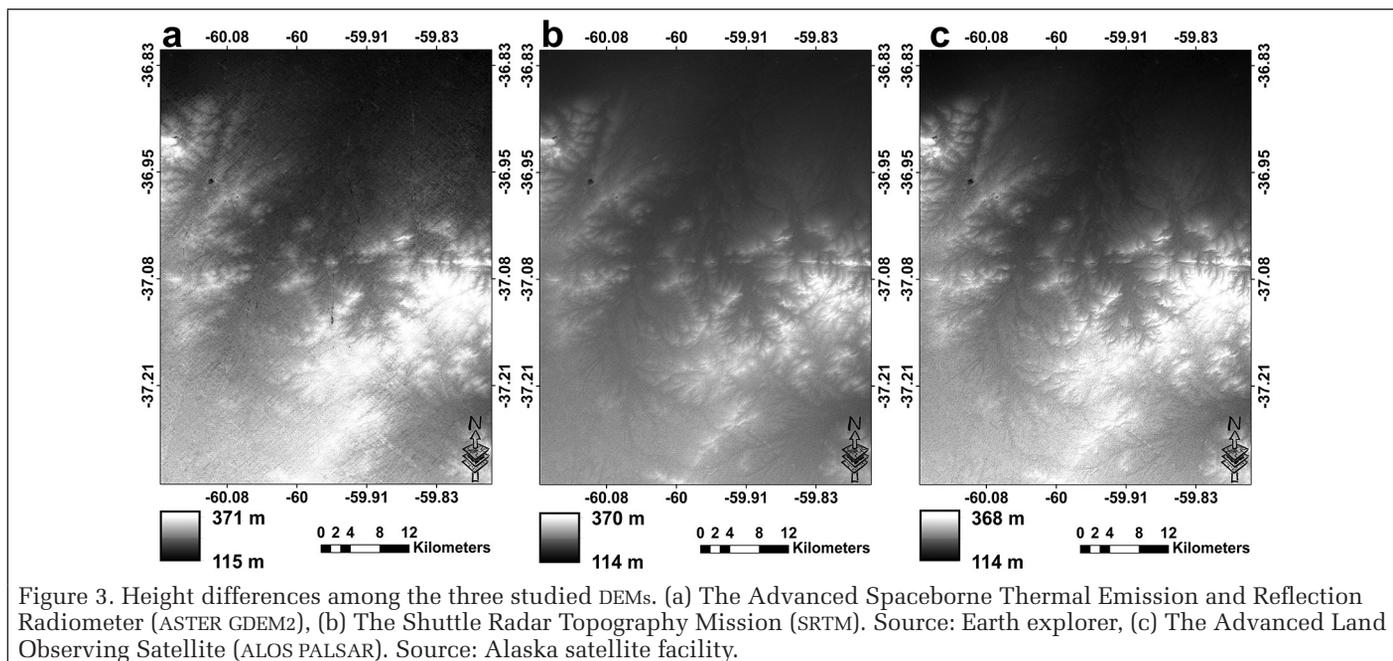


Figure 3. Height differences among the three studied DEMs. (a) The Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER GDEM2), (b) The Shuttle Radar Topography Mission (SRTM). Source: Earth explorer, (c) The Advanced Land Observing Satellite (ALOS PALSAR). Source: Alaska satellite facility.

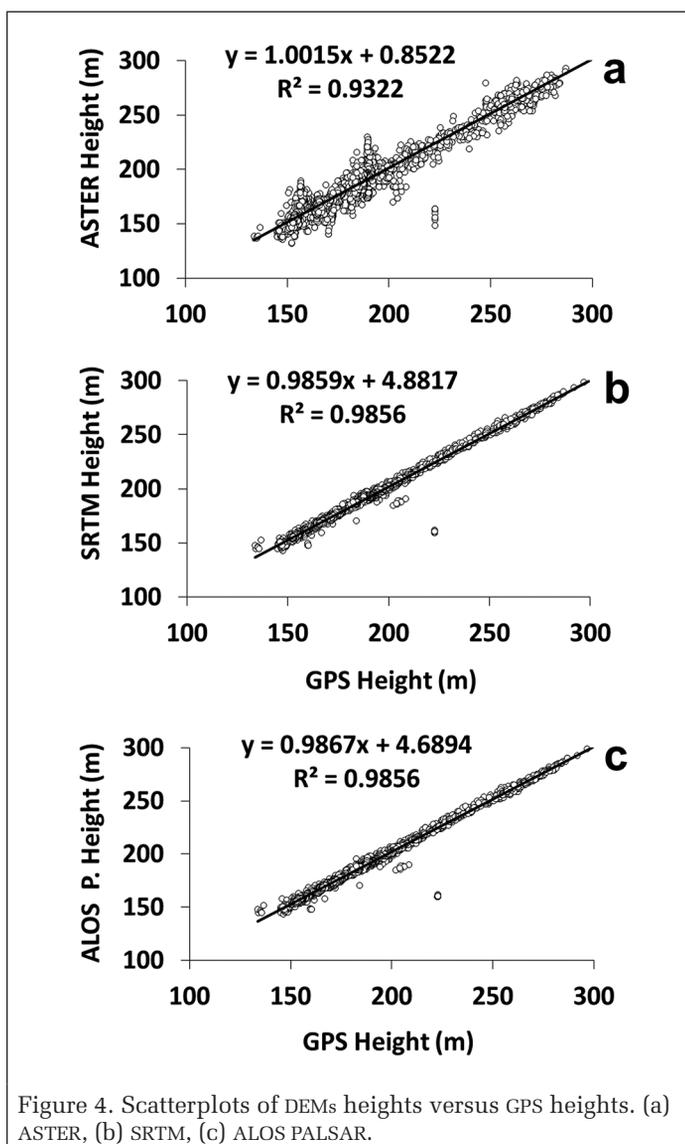


Figure 4. Scatterplots of DEMs heights versus GPS heights. (a) ASTER, (b) SRTM, (c) ALOS PALSAR.

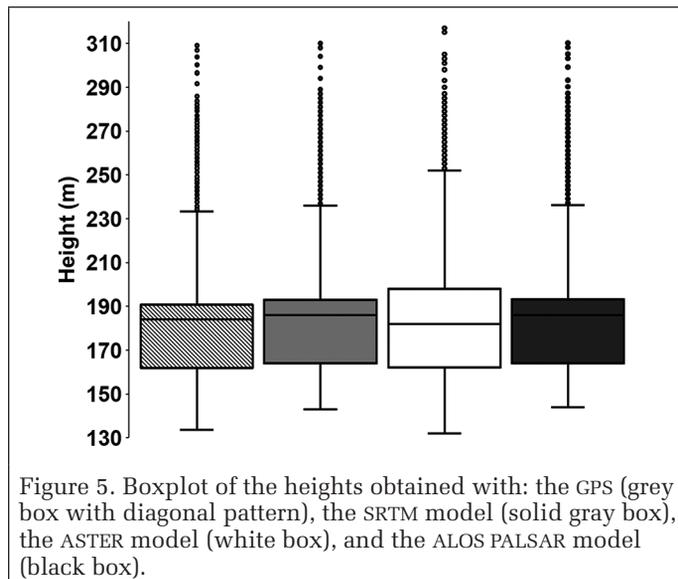


Figure 5. Boxplot of the heights obtained with: the GPS (grey box with diagonal pattern), the SRTM model (solid grey box), the ASTER model (white box), and the ALOS PALSAR model (black box).

Analysis of Topography

The topography of plain zones is one of the main factors affecting DEMs vertical accuracy due to the low morphometric potential of these areas. To evaluate topography, the DEMs were divided into four different elevation bands (<150 m, 150–200 m, 200–250 m, and >250 m) and the slope was analyzed considering two ranges ($\leq 3\%$, $> 3\%$).

Figure 6a shows the obtained areas within each elevation band using the ASTER, the SRTM and the ALOS PALSAR models. The greatest difference was in between the 150–250 m range.

Vertical accuracy regarding the relative area is shown in Figure 6b, which indicates that 10% of the area had <5-m errors, 40% of the area had 4.5–8.5 m errors, 85% of the area had 3.3–7.7 m errors.

Figure 6c presents the RMSE of the elevation grouped by elevation band for each of the studied DEMs (ASTER, SRTM and ALOS PALSAR). A major error was detected in the three models for the 150–200 m elevation band, thus causing an increase in the uncertainty of the measurements within this height range. A greater error range was found for the ASTER model with a ± 5.0 –8.5 m RMSE. The RMSE for SRTM was ± 2.60 –4.5 m whereas for the ALOS PALSAR RMSE was ± 2.5 –4.5 m. Regarding

the topography representation, SRTM and ALOS PALSAR offered greater precision both in the upper and lower zones compared to ASTER.

In addition, the vertical accuracy among ASTER, SRTM and ALOS PALSAR was evaluated in terms of slope (Figure 6d), showing that the SRTM and the ALOS PALSAR models presented better fit slopes (RMSE ± 4.4 m for both) compared to ASTER (RMSE ± 9.1 m).

Geomorphological Analysis

The analysis of the geomorphs focused on the deflation hollows. Due to the characteristics of the study area they were grouped into two categories, one corresponding to the high zone of the study area with $>3\%$ -slopes and another corresponding to the low zone of the study area with $<3\%$ -slopes (Figure 7). From the three tested models, the ALOS PALSAR best represented the shape of the deflation hollows. The RMSE for this model was ± 1.8 m in the high zone and ± 2.3 m in the low zone, with a STD: $\pm 1-1.25$ m. The RMSE values for SRTM were ± 2.8 and ± 3.4 m in the high and low zones, respectively, and a STD: $\pm 1.58-1.8$ m. The ASTER model did not delineate correctly the deflation hollows. RMSE values were ± 3.4 and ± 3.5 m for the high and the low zones, respectively, with STD being $\pm 1.9-2.2$ m.

Analysis of DEM Response to the Presence of Civil Structures

Due to the low morphometry of the study area, main and secondary roads are raised onto embankments, so that they can be functional even in periods of water excess. These embankments form a barrier affecting surface flow and lead to a change in the direction and amount of water moving over the surface.

The accuracy of the three DEMs was evaluated in terms of how effectively they represented these structures. As we did for the analysis of the geomorphs, the study area was divided into a high and a low zone. We found out that the SRTM model best represented the main and secondary roads as observed in Figure 8. For the main roads, the RMSE for SRTM was $\pm 1.9-6$ m (STD: $\pm 4.4-10.7$ m) whereas for the ALOS PALSAR and ASTER RMSE values were $\pm 2.3-6.2$ m (STD: $\pm 4.6-10.8$ m) and $\pm 6.7-12$ m (STD: $\pm 5.2-11.8$ m), respectively. For the secondary roads the SRTM showed a RMSE of ± 2.6 m, whereas for the ALOS PALSAR and ASTER RMSE values were ± 2.75 m and ± 9.7 m, respectively.

Evaluation of Basin Delimitation and Drainage Network

Basin delimitation and drainage network analysis in the study area were carried out using the Archydro extension of ArcGIS as described in the methods section the location coordinates at the watershed outlet point were -36.83° S and -59.89° W.

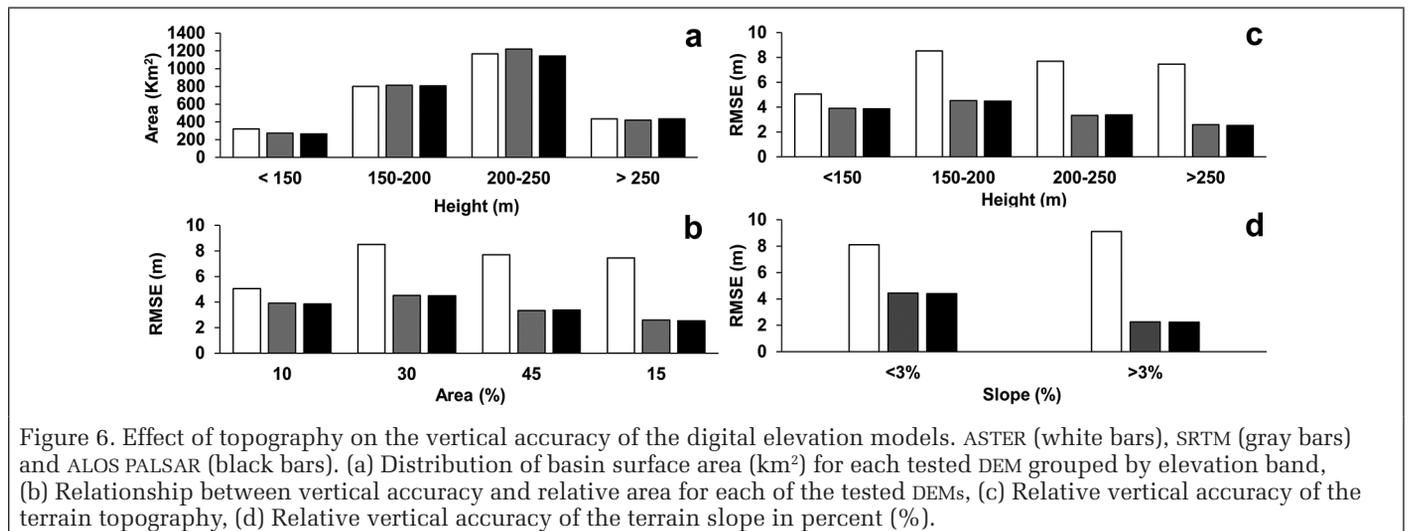


Figure 6. Effect of topography on the vertical accuracy of the digital elevation models. ASTER (white bars), SRTM (gray bars) and ALOS PALSAR (black bars). (a) Distribution of basin surface area (km²) for each tested DEM grouped by elevation band, (b) Relationship between vertical accuracy and relative area for each of the tested DEMs, (c) Relative vertical accuracy of the terrain topography, (d) Relative vertical accuracy of the terrain slope in percent (%).

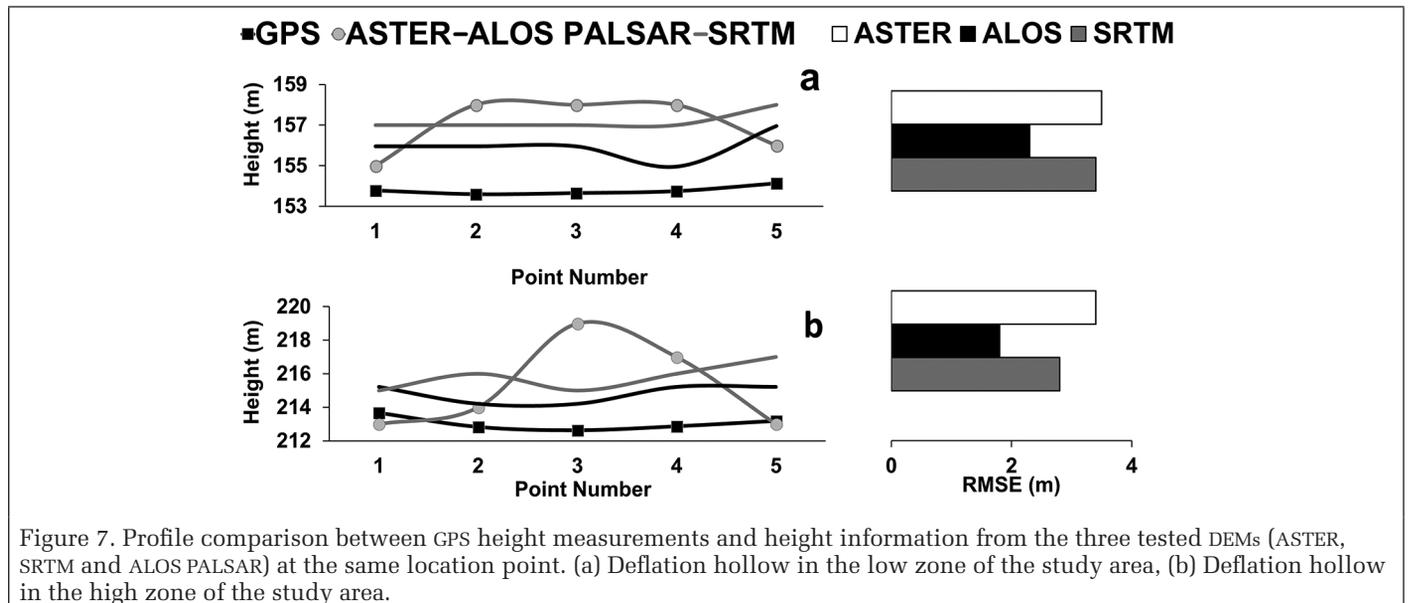


Figure 7. Profile comparison between GPS height measurements and height information from the three tested DEMs (ASTER, SRTM and ALOS PALSAR) at the same location point. (a) Deflation hollow in the low zone of the study area, (b) Deflation hollow in the high zone of the study area.

The drainage network was generated using a threshold surface area of 2100 ha. (i.e., only areas above this threshold value were considered). Basin delimitation was better represented by SRTM and ALOS PALSAR than by ASTER. The former DEMs produced drainage networks which fitted more closely to the drainage network obtained from digitized topographic maps at 1:100 000 scale (Figure 9), both for the upper and lower part of the basin.

The ASTER was the worse of the three DEMs in correctly delimiting the drainage network. This DEM produced inconsistencies especially in for the upper part of the basin (Figure 9). This inconsistencies could be attributed to the photogrammetric DEM overestimating the lower order streams as suggested by Thomas and Prasanna Kumar (2015) and Dass and Pardeshi (2018). These errors lead to the obtention of different runoff surfaces areas among DEMs (Table 1), translating into future errors when performing hydrological modeling either for distributed or semi-distributed models, due to the fact that the surface-subsurface flow and concentration time are misestimated.

As for the drainage network, the accuracy of each DEM was analyzed in terms of how well the shape of the stream channel in the plain zone of the basin was represented. The three DEMs had difficulties representing the stream channel in the study area (Figure 10), although the SRTM and ALOS PALSAR models better accounted for the variation in the topography and better resembled the shape of the channel compared to the ASTER model. The following were the RMSE and STD for each of the tested DEMs: SRTM (RMSE: ± 2.2 – 6.8 m, STD: ± 1.3 – 3.6 m), ASTER (RMSE: ± 3.3 – 6.1 m, STD: ± 2.3 – 3.3 m) and ALOS PALSAR (RMSE: ± 2.5 – 6.8 m, STD: ± 1.4 – 3.7 m). These errors were greater in the lower zone of the study area.

Interpolation Methods

When analyzing the previous results, the models that better represented the topography of the study area were SRTM and ALOS PALSAR. From these two, SRTM was the best at representing the topography in flat areas with a lower density of sampling points. Hence, this was the selected DEM for the analysis of the different interpolation methods. The raw SRTM model in raster format was transformed into its equivalent vector format by means of points, (one point every 30 m). These points were resampled using the different interpolators in order to improve the accuracy of the DEM in the plain zone.

We evaluated the efficiency of five interpolators (i.e., inverse distance weighting, natural neighbor, spline, kriging and ANUDEM). The sensitivity of the terrain to various interpolators was also analyzed in relation to the slope and elevation bands in the study area. The accuracy of the generated model of terrain depends on the interpolation mechanism adopted, and therefore it is necessary to investigate the comparative performance of the different methodological approaches.

The method of interpolation that showed more discrepancies in representing the terrain topography (elevation and slope) was the spline method with an error of ± 2.72 – 4.61 m. The interpolation methods that better represented the elevation bands in the study area (Figure 11a) were, in decreasing order, the inverse distance weighting (RMSE: ± 1.7 – 3.2 m), the ANUDEM (RMSE: ± 2 – 3.3 m) and natural neighbor (RMSE: ± 1.8 – 3.4 m). The kriging method showed a ± 2.7 – 3.9 m RMSE.

Figure 11b shows that the interpolation methods that better represented the slope in the study area were the following, in order of importance: inverse distance weighting (RMSE: ± 1.5 – 3.1 m), ANUDEM and natural neighbor (RMSE: ± 1.6 – 3.2 m for both) and the kriging method (RMSE: ± 2.2 – 3.9 m).

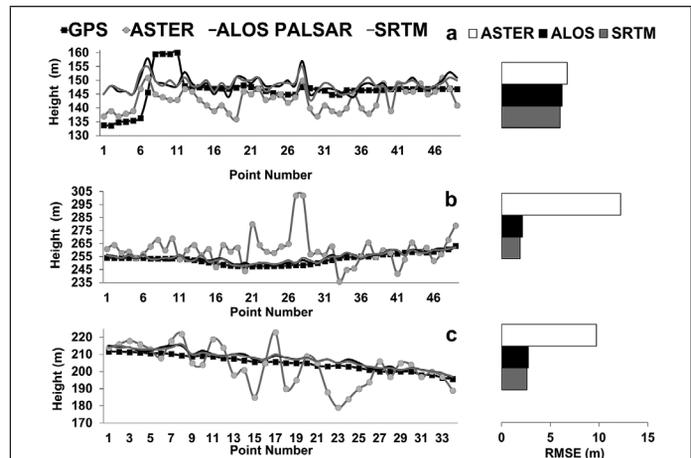


Figure 8. Profile comparison between GPS measured heights and heights provided by each of the three DEMs (ASTER, SRTM and ALOS PALSAR) at locations presenting civil structures. (a) highway at the low zone of the study area, (b) highway at the high zone of the study area, (c) secondary road.

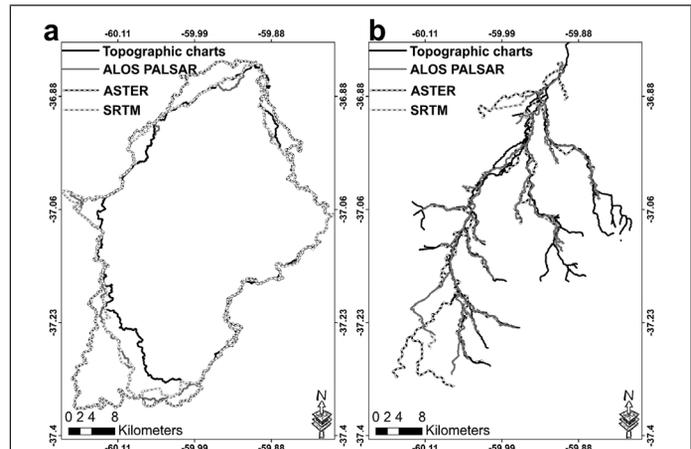


Figure 9. (a) Basin delimitation, (b) Drainage network, for Del Azul upper creek basin obtained from different. Sources: topographic charts, ASTER, SRTM and ALOS PALSAR.

Discussion

Our study contributes to improving the information provided by DEMs for plain areas and presents an appropriate methodology for increasing height accuracy and for evaluating the effect of the uncertainties in the currently available models. We evaluated the performance of these models to predict in a realistic way the topography, surface area, slope and flow processes in terms of direction and accumulation, so that they become suitable for use in hydrological modeling. This is to satisfy the need of accurate information for water balance due to it is very sensitive to the topographic attributes, geomorphological and civil structures present in the study area.

Regarding height accuracy, the radar models (i.e., SRTM and ALOS PALSAR) represented more accurately the topography in plain areas. This is because the sensors with which these

Table 1. Differences between areas of the basin delimited by different means.

Source	Area of the Basin (km ²)	Drainage Length (km)	Drainage Density (km/km ²)
ASTER	1213	252	4.81
SRTM	1050	214	4.9
ALOS PALSAR	1046	220	4.75
Topographic charts	982	312	3.2

models work are more accurate than the one for ASTER, capturing with more detail the Earth's surface (De Oliveira and Paradella 2008; Frey and Paul 2012; Frey *et al.* 2012; De Oliveira and De Fátima 2012; Rossetti 2012; Eric *et al.* 2014; Rexer and Hirt 2014; Tan *et al.* 2015; Jarihani *et al.* 2015; Das *et al.* 2015).

One of the drawbacks of DEMs use in plain zones is the poor representation of channels, civil structures and geomorphology. This is because there are <1 m differences in land height and even this difference is small, it plays a significant role in the movement of surface and subsurface water in this type of areas. The SRTM and the ALOS PALSAR were the DEMs that better represented the topography of the study area, unlike ASTER, which showed serious inconsistencies at representing the geomorphology and civil structures. Indeed, Eric *et al.* (2014) stated that the ASTER model had problems in delineating morphological units.

Regarding the behavior of the interpolators, we found out that these were sensitive to the attributes of the terrain. The inverse distance weighting method demonstrated the best accuracy compared to the other interpolation methods. The generated DEM for the study area based on the SRTM with IDW interpolated data showed a 33% error reduction using, the ANUDEM method, a 27% reduction using the natural neighbor method and a 10% reduction using the kriging method whereas the spline method increased the error by 4% compared to the raw SRTM DEM.

Our study suggests that the most suitable interpolation methods to adequate DEMs for use in plain zones are the inverse distance weighting and ANUDEM. This statement agrees with other published studies such as Aguilar *et al.* (2005), Reuter *et al.* (2007), Paredes *et al.* (2007) and Guo *et al.* (2010).

Our findings showed that topographic variability significantly influences DEMs accuracy and the degree of detail they provide. However, future research is needed to further explore the kriging interpolation method, the semivariogram analysis and the effect of the sampling density, all of which could improve the representation of topographic variability and increase the vertical accuracy.

Regarding the representation of the flow direction and water accumulation, in plain zones, the SRTM model generated a more realistic drainage network latter translated into a better delineated basin (as compared with the basin boundaries obtained using the topographic chart). If we use this product for hydrological modeling it is important to bear in mind that the choice and correction of the DEM will influence the calculation of surface runoff and the flood response time. According to Thomas and Prasanna Kumar (2015) and Dass and Pardeshi (2018), a photogrammetric DEM such as ASTER has inconsistencies in the representation of the drainage networks.

The needs of having readily available elevation data at a detailed scale, affordable and easy to use, has stimulated the development of several satellite platforms designed to generate this type of Earth's surface information. However, although precision ranges are generally well described for each platform, various authors recommend that DEM validations should be performed locally to arrive to more accurate results in order to take better decisions.

Our purpose with this study is to provide DEM users with a set of tools to analyze and evaluate the sensitivity and the uncertainties associated using different topographic parameters.

Conclusions

As seen through this paper, DEMs accuracy to represent terrain morphology, geomorphology and the presence of civil structures in plain areas differ among models. The selection of which DEM to use will have great influence on the modeled water flow patterns. From the three tested DEMs, the SRTM and the ALOS PALSAR were the ones that best represented these

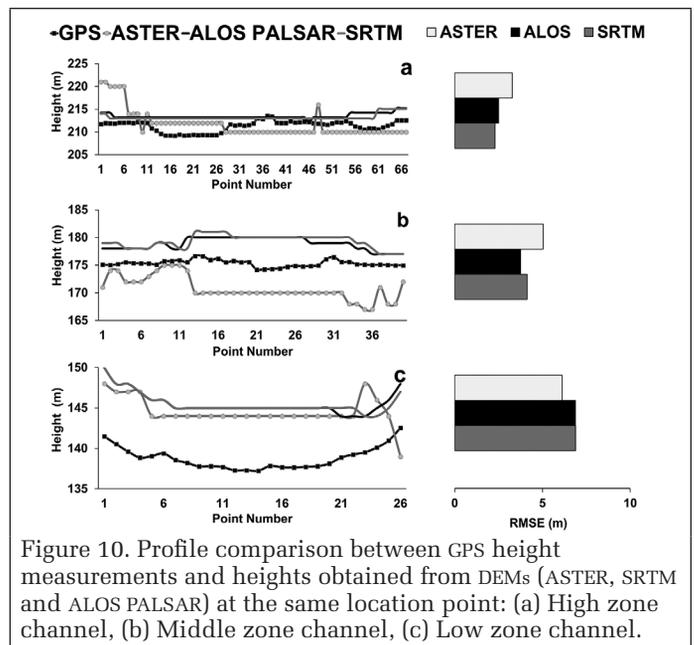


Figure 10. Profile comparison between GPS height measurements and heights obtained from DEMs (ASTER, SRTM and ALOS PALSAR) at the same location point: (a) High zone channel, (b) Middle zone channel, (c) Low zone channel.

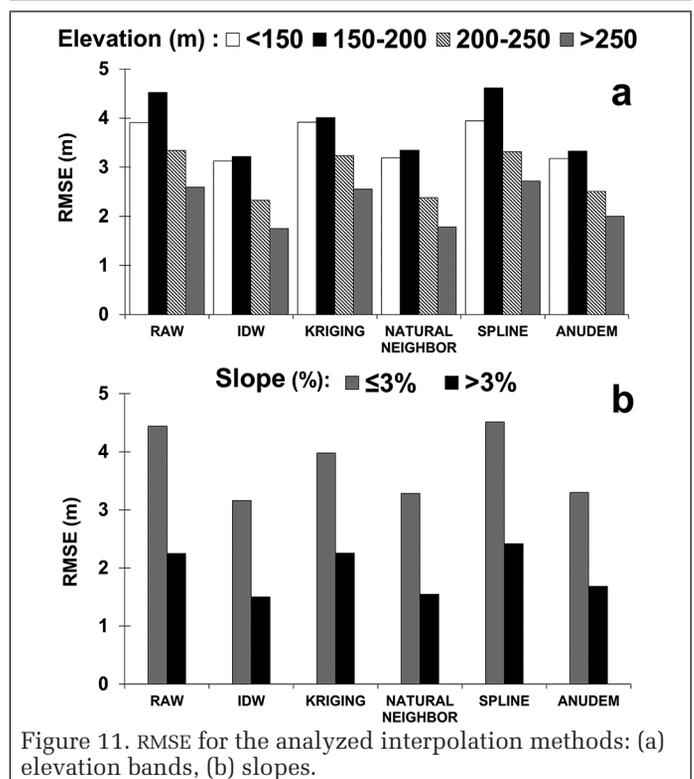


Figure 11. RMSE for the analyzed interpolation methods: (a) elevation bands, (b) slopes.

topographic variations, with error values of 4.4 m for both. In contrast, the ASTER model presented inconsistencies in delimiting landscape units, with error values above 8 m.

The SRTM and the ALOS PALSAR DEMs showed a better performance in the representation of the morphology for the plain area of the basin whereas the ASTER DEM overestimated the low-stream-order-drainage network and thus provided an inaccurate delimitation of the basin. Therefore, special attention should be taken when choosing the particular DEM for hydrological modeling in plain areas since this choice will surely influence the response time and amount of water moving over the surface.

When analyzing and visualizing the vertical error distribution in plain zones, the spline method is not recommended for elevation data interpolation since this method showed the

lowest statistical precision among the analyzed methods. The IDW and ANUDEM methods increased the accuracy of the model in approximately 33% from a raw SRTM (RMSE: ± 4.44 m and ± 3.2 m for the non-corrected and corrected models, respectively).

Summarizing, when hydrological studies are carried out in plain zones, it is necessary to know beforehand that the SRTM, the ALOS PALSAR and the ASTER DEMs do not correctly reflect the water flow patterns. The associated errors must be corrected by adjusting the heights in the models with the aid of validated information sources such as topographical charts, satellite images, interpolations and vectors based on measurements performed in the field with differential GPS, particularly in roads, channels and depressions. All these considerations should at least be taken into account in order to improve the accuracy of the DEMs for use in plain areas.

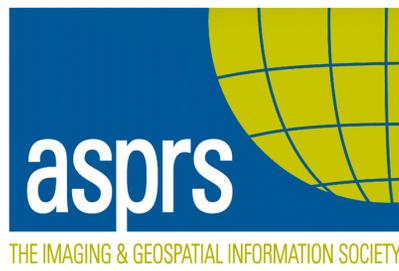
Acknowledgments

The authors would like to thank the Instituto de Hidrología de Llanuras, Dr. Eduardo Usunoff (IHLLA) and the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) for funding this study.

References

- Abrams, M. and S. Hook. 2002. ASTER User Handbook Version 2. Jet Propulsion, 1–135.
- Abrams, M., H. Tsu, G. Hulley, K. Iwao, D. Pieri, T. Cudahy, and J. Kargel. 2015. The Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) after fifteen years: Review of global products. *International Journal of Applied Earth Observation and Geoinformation* 38: 292–301.
- Aguilar, F. J., F. Agüera, M. a. Aguilar and F. Carvajal. 2005. Effects of terrain morphology, sampling density, and interpolation methods on grid DEM accuracy. *Photogrammetric Engineering Remote Sensing* 71 (7): 805–816.
- Arun, P. V. 2013. A comparative analysis of different DEM interpolation methods. *The Egyptian Journal of Remote Sensing and Space Science* 16 (2): 133–139.
- Barnes, R., C. Lehman and D. Mulla, 2014. An efficient assignment of drainage direction over flat surfaces in raster digital elevation models. *Computers Geosciences* 62:128–135.
- Bartier, P. M. and C. P. Keller. 1996. Multivariate interpolation to incorporate thematic surface data using inverse distance weighting (IDW). *Computers Geosciences* 22 (7): 795–799.
- Callow, J. N., K. P. Van Niel and G. S. Boggs. 2007. How does modifying a DEM to reflect known hydrology affect subsequent terrain analysis? *Journal of Hydrology* 332 (1–2):30–39.
- Carrascal Leal, C. C., G. A. Denegri and M. I. Delgado. 2013. Costos mínimos de compensación y cuantificación de la oferta hídrica en la cuenca alta del río Sauce Grande, Argentina. *Investigaciones Geográficas* 0 (80):55–70.
- Chaubey, I., A. S. Cotter, T. A. Costello and T. S. Soerens. 2005. Effect of DEM data resolution on SWAT output. 628:621–628.
- Chen, Y., J. P. Wilson, Q. Zhu and Q. Zhou. (2012). Comparison of drainage-constrained methods for DEM generalization. *Computers Geosciences* 48:41–49.
- Childs, B. C. 2004. Interpolation surfaces in ArcGIS spatial analyst. *ArcUser*, July–September 2004: 32–35.
- Chu, T. and K. E. Lindenschmidt. 2017. Comparison and validation of digital elevation models derived from InSAR for a flat inland delta in the high latitudes of Northern Canada. *Canadian Journal of Remote Sensing* 43 (2):109–123.
- Das, A., R. Agrawal and S. Mohan. 2015. Topographic correction of ALOS-PALSAR images using InSAR-derived DEM. *Geocarto International* 30 (2):145–153.
- Das, S. and S. D. Pardeshi. 2018. Comparative analysis of lineaments extracted from Cartosat, SRTM and ASTER DEM: A study based on four watersheds in Konkan region, India. *Spatial Information Research* 26 (1):47-57.
- De Oliveira Andrades Filho, C. and D. De Fátima Rossetti. 2012. Effectiveness of SRTM and ALOS-PALSAR data for identifying morphostructural lineaments in northeastern Brazil. *International Journal of Remote Sensing* 33 (4):1058–1077.
- De Oliveira, C. G. and W. R. Paradella. 2008. An assessment of the altimetric information derived from spaceborne SAR (RADARSAT-1, SRTM3) and optical (ASTER) data for cartographic application in the Amazon region. *Sensors* 8 (6):3819–3829.
- Easton, Z. M., D. R. Fuka, M. T. Walter, D. M. Cowan, E. M. Schneiderman and T. S. Steenhuis. 2008. Re-conceptualizing the soil and water assessment tool (SWAT) model to predict runoff from variable source areas. *Journal of Hydrology* 348 (3–4):279–291.
- Eric, Z., W. Petrus, D. Clercq and A. V. Niekerk. 2014. An evaluation of digital elevation models (DEMs) for delineating land components. *Geoderma*, 213 :312–319.
- Frey, H., F. Paul and T. Strozzi. 2012. Compilation of a glacier inventory for the western Himalayas from satellite data: Methods, challenges, and results. *Remote Sensing of Environment*, 124 :832–843.
- Gallant, J. C. and T. I. Dowling. 2003. A multiresolution index of valley bottom flatness for mapping depositional areas. *Water Resources Research* 39 (12).
- Gao, X., X. Zhao, P. Wu, L. Brocca and B. Zhang. 2016. Effects of large gullies on catchment-scale soil moisture spatial behaviors: A case study on the Loess Plateau of China. *Geoderma* 261:1–10.
- Gesch, D. B., J. Muller and T. G. Farr. 2006. The shuttle radar topography mission data validation and applications. *Photogrammetric Engineering Remote Sensing* 72:233–235.
- Gichamo, T. Z., I. Popescu, A. Jonoski and D. Solomatine. 2012. River cross-section extraction from the ASTER global DEM for flood modeling. *Environmental Modelling Software* 31:37–46.
- Guevara, C. 2015. Una metodología para el manejo integral de extremos hídricos en una cuenca rural en zona de llanura. caso de estudio: cuenca arroyo Santa Catalina, provincia de Buenos Aires. *Tesis de Maestría en Ecohidrología, Universidad Nacional de la Plata*, 143.
- Guo, Q., W. Li, H. Yu and O. Alvarez. 2010. Effects of topographic variability and Lidar sampling density on several DEM interpolation methods. *Photogrammetric Engineering Remote Sensing* 76:701–712.
- Hidayat, H., A. J. Teuling, B. Vermeulen, M. Taufik, K. Kastner, T. J. Geertsema and A. J. F. Hoitink. 2017. Hydrology of inland tropical lowlands: The Kapuas and Mahakam wetlands. *Hydrology and Earth System Sciences* 21:2579–2594.
- Hutchinson, M. F. 1989. A new procedure for gridding elevation and stream line data with automatic removal of spurious pits. *Journal of Hydrology* 106 (3–4):211–232.
- Hutchinson, M., T. Xu and J. Stein. 2011. Recent progress in the ANUDEM elevation gridding procedure. *Geomorphometry* 19–22.
- Jarihani, A., J. N. Callow, T. R. McVicar, T. G. Van Niel and J. R. Larsen. 2015. Satellite-derived Digital Elevation Model (DEM) selection, preparation and correction for hydrodynamic modelling in large, low-gradient and data-sparse catchments. *Journal of Hydrology* 524:489–506.
- Kellndorfer, J., W. Walker, L. Pierce, C. Dobson, J. A. Fites, C. Hunsaker and M. Clutter. 2004. Vegetation height estimation from shuttle radar topography mission and national elevation datasets. *Remote Sensing of Environment* 93 (3):339–358.
- Kimura, H. and N. Ito. 2000. ALOS/PALSAR: The Japanese second-generation spaceborne SAR and its applications. In *Proceedings of SPIE*. 4152:110–119.
- Lemoine, F. G., S. C. Kenyon, J. K. Factor, R. G. Trimmer, N. K. Pavlis, D. S. Chinn and T. R. Olson. 1998. The development of the NASA GSFC and NIMA joint geopotential model. *Gravity, Geoid and Marine Geodesy* 117:461–469.
- Lin, S., C. Jing, N. A. Coles, V. Chaplot and N. J. Moore. 2013. Evaluating DEM source and resolution uncertainties in the Soil and Water Assessment Tool. *Stochastic Environmental Research and Risk Assessment* 27 (1):209-221.

- Ludwig, R. and P. Schneider. 2006. Validation of digital elevation models from SRTM X-SAR for applications in hydrologic modeling. *ISPRS Journal of Photogrammetry and Remote Sensing* 60 (5):339–358.
- Martz, L. W. and E. Jong. 1988. CATCH: A FORTRAN program for measuring catchment area from digital elevation models. *Computers Geosciences* 14 (5):627–640.
- Moore, I. D., R. B. Grayson and R. Ladson. 1991. Digital terrain modeling: A review of hydrological geomorphological and biological applications. *hydrological processes*. 5 (1):3–30.
- Mukherjee, S., P. K. Joshi, S. Mukherjee, A. Ghosh, R. D. Garg and A. Mukhopadhyay. 2013. Evaluation of vertical accuracy of open source Digital Elevation Model (DEM). *International Journal of Applied Earth Observation and Geoinformation*, 21:205–217.
- Pachri, H., Y. Mitani, H. Ikemi, I. Djameluddin and A. Morita. 2013. Development of water management modeling by using GIS in Chirchik River Basin, Uzbekistan. *Procedia Earth and Planetary Science* 6:169–176.
- Paredes, C., W. Salinas, X. Becerra and S. Jiménez. 2013. Evaluación y comparación de métodos de interpolación determinísticos y probabilísticos para la generación de modelos digitales de elevación. *Investigaciones Geográficas* 82:118–130.
- Pavlis, N. K., S. A. Holmes, S. C. Kenyon and J. K. Factor. 2012. The development and evaluation of the Earth Gravitational Model 2008 (EGM2008). *Journal of Geophysical Research: Solid Earth* 117 (B4):1–38.
- Pontes, P. R. M., F. M. Fan, A. S. Fleischmann, R.C.D. de Paiva, D. C. Buarque, V. A. Siqueira and W. Collischonn. 2017. MGB-IPH model for hydrological and hydraulic simulation of large floodplain river systems coupled with open source GIS. *Environmental Modelling Software* 94:1–20.
- Rabus, B., M. Eineder, A. Roth and R. Bamler. 2003. The shuttle radar topography mission—a new class of digital elevation models acquired by spaceborne radar. *ISPRS Journal of Photogrammetry and Remote Sensing* 57 (4):241–262.
- Reuter, H. I., A. Nelson and A. Jarvis. 2007. An evaluation of void filling interpolation methods for SRTM data. *International Journal of Geographical Information Science* 21 (9):983–1008.
- Rexer, M. and C. Hirt. 2014. Comparison of free high-resolution digital elevation data sets (ASTER GDEM2, SRTM v2.1/v4.1) and validation against accurate heights from the Australian National Gravity Database. *Australian Journal of Earth Sciences* 61 (2):213–226.
- Robinson, N., J. Regetz and R. P. Guralnick. 2014. EarthEnv-DEM90: A nearly-global, void-free, multi-scale smoothed, 90m digital elevation model from fused ASTER and SRTM data. *ISPRS Journal of Photogrammetry and Remote Sensing* 87:57–67.
- Rosenqvist, A., M. Shimada, N. Ito and M. Watanabe. 2007. ALOS PALSAR: A pathfinder mission for global-scale monitoring of the environment. *IEEE Transactions on Geoscience and Remote Sensing* 45 (11):3307–3316.
- Sanders, B. F. 2007. Evaluation of on-line DEMs for flood inundation modeling. *Advances in Water Resources* 30 (8):1831–1843.
- Scioli, C. 2009. Modelación del escurrimiento superficial en áreas de llanura: Implementación y calibración de un modelo distribuido de grilla. *Tesis de Maestría en Hidrología en Llanuras, Universidad de Rosario* 117.
- Sharma, A. and K. N. Tiwari. 2014. A comparative appraisal of hydrological behavior of SRTM DEM at catchment level. *Journal of Hydrology* 519:1394–1404.
- Shortridge, A. and J. Messina. 2011. Spatial structure and landscape associations of SRTM error. *Remote Sensing of Environment* 115 (6):1576–1587.
- Sibson, R. 1981. A brief description of natural neighbour interpolation. *Interpreting Multivariate Data* 21–36.
- Sun, G., K. Ranson, V. Kharuk and K. Kovacs. 2003. Validation of surface height from shuttle radar topography mission using shuttle laser altimeter. *Remote Sensing of Environment* 88 (4):401–411.
- Tachikawa, T., M. Hato, M. Kaku and A. Iwasaki. 2011. Characteristics of ASTER GDEM version 2. pp. 3657–3660 in *Proceedings Geoscience and Remote Sensing Symposium (IGARSS)*.
- Tan, M. L., D. L. Ficklin, B. Dixon, A. L. Ibrahim, Z. Yusop and V. Chaplot. 2015. Impacts of DEM resolution, source, and resampling technique on SWAT-simulated streamflow. *Applied Geography* 63:357–368.
- Tarboton, D. G. 1997. A new method for the determination of flow directions and upslope areas in grid digital elevation models. *Water Resources Research* 33 (2):309–319.
- Tarekegn, T. H., A. T. Haile, T. Rientjes, P. Reggiani and D. Alkema. 2010. Assessment of an ASTER-generated DEM for 2D hydrodynamic flood modeling. *International Journal of Applied Earth Observation and Geoinformation* 12 (6):457–465.
- Thomas, J. and V. Prasannakumar. 2015. Comparison of basin morphometry derived from topographic maps, ASTER and SRTM DEMs: An example from Kerala, India. *Geocarto International* 30 (3):346–364.
- Vaze, J., J. Teng and G. Spencer. 2010. Impact of DEM accuracy and resolution on topographic indices. *Environmental Modelling Software* 25 (10):1086–1098.
- Wahba, G. and J. Wendelberger. 1980. Some new mathematical methods for variational objective analysis using splines and cross validation. *Monthly Weather Review*.
- Wechsler, S. P. and C. N. Kroll. 2006. Quantifying DEM uncertainty and its effect on topographic parameters. *Photogrammetric Engineering Remote Sensing*, 72 (9):1081–1090.
- Wilson, J. P. 2012. Digital terrain modeling. *Geomorphology* 137 (1):107–121.
- Wilson, J. P. and J. C. Gallant. 2000. Digital terrain analysis. *Terrain Analysis: Principles and Applications* 6 (12):1–27.
- Wurl, J., C. N. Martínez García and M. Á. Imaz Lamadrid. 2014. Caracterización del peligro por inundaciones en el oasis La Purísima, Baja California Sur, México. *Investigaciones Geográficas, Boletín Del Instituto de Geografía* 87:76–87.
- Yan, K., A. Tarpanelli, G. Balint, T. Moramarco and G. Baldassarre. 2014. Exploring the Potential of SRTM Topography and Radar Altimetry to Support Flood Propagation Modeling: Danube Case Study. *Journal of Hydrologic Engineering* 20 (2):1–7.
- Yang, X. and T. Hodler. 2013. Visual and statistical comparisons of surface modeling techniques for point-based environmental data. *Cartography and Geographic Information Science* 38 (1):20–35.
- Zabala, M. E., M. Manzano and L. Vives. 2015. The origin of groundwater composition in the Pampeano Aquifer underlying the Del Azul Creek basin, Argentina. *Science of the Total Environment* 518: 168–188.
- Zhang, T., X. Xu and S. Xu. 2015. Method of establishing an underwater digital elevation terrain based on kriging interpolation. *Measurement* 63:287–298.



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

An Evaluation of Reflectance Calibration Methods for UAV Spectral Imagery

Jarrod Edwards, John Anderson, William Shuart, and Jason Woolard

Abstract

Spectral imagery using micro-unmanned aerial vehicles is rapidly advancing. This study compared reflectance calibration methods for imagery acquired using the Parrot Sequoia imager, a commercial multispectral sensor package. For the study, two orthomosaics were calibrated using 1) a manufacturer-suggested AIRINOV standard correction using PIX4D software and 2) the Empirical Line Calibration (ELC) method using ground radiometric data on specific in-scene targets. Both scenes were analyzed for target spectral agreement by ground radiometric survey. Regression analysis demonstrated more favorable target correlation for the ELC imagery than the AIRINOV-calibrated imagery, with Root Mean Square Error (RMSE) analysis supporting these results. Finally, classification maps were produced between the data sets. Error analysis resulted in an overall accuracy of 24% for the AIRINOV map compared to ELC-based truth data, with a considerable number of pixels associated with brighter targets unclassified. These results demonstrate the need for standardized calibration procedures in the spectral correction of small-format remote sensor data.

Key words: micro-UAS, UAV, drone, remote sensing, spectral reflectance, empirical line calibration, classification, ground radiometry, error analysis, multispectral.

Introduction

The data produced from air and satellite-based remote sensor systems has dominated the geospatial sciences for many decades with a rich variety of spatial and spectral imagery formats. These data have contributed significantly to the social, natural and physical science disciplines with products of varied quality and utility for a variety of mapping applications (Haboudane *et al.* 2004; Mumby *et al.* 1998; Cowley *et al.* 2018). Many of these data are also largely verifiable and collected using standardized, well-documented correction procedures that an end-user can make reference to when exploiting and certifying the data or resulting products (Price 1987; Cocks *et al.* 1998). For example, Thematic Mapper is available in various radiometrically corrected configurations and post-processed product levels that are rigorously researched and provide the user an idea of the quality of the data with relation to spectral fidelity (Thome *et al.* 1993; Chander *et al.* 2009; N. Mishra *et al.* 2016; Chander and Markham 2003; Chander *et al.* 2007). In addition, airborne multi- and hyperspectral data are provided in formats that

typically report calibration methods that include atmospheric correction routines and reflectance calibration procedures (Richter and Schläpfer 2002; Eismann 2012; Gao *et al.* 2009). Until recently, data integrity has been discussed and presented as a verifiable and traceable part of many of the systems remote sensing has depended upon (Schaeppman-Strub 2006). However, the last decade has seen the explosive development of small-format imagers paired with unmanned aerial systems (UAS) platforms that have offered greater flexibility and more-than-competitive capabilities in spatial, spectral, and temporal resolution (Elias 2012; Colomina and Molina 2014). These imaging systems can include (non-metric) compact to sub-compact cameras of various quality that produce frame imagery collected using a combination of on-board global positioning system (GPS) and inertial measurement system triggers that are post processed using analytical photogrammetry. The benefits and flexibility UAS offers includes size, temporal exploitation, and an ever-widening availability of sensor payloads. But while the size and weight of many systems are convenient for transport and deployment, it does not typically translate into a stable sensor platform in the air. In fact, many systems are highly unstable in flight serving to introduce post-processed spatial and radiometric errors associated with changing viewing geometry and photogrammetric exterior orientation issues. These changes often include topography and sun angle issues manifested as bidirectional reflectance distribution function (BRDF) variations in the individual images comprising a final image ortho-mosaic (Burkart *et al.* 2015; Stark *et al.* 2016). Both the potential platform stability issues and a shortage of research on the spectral fidelity of these new compact sensors beg for validation standards for drone-based electro-optic (EO) data similar to those described by Justice *et al.* (2010) for satellite sensor products.

There are a number of UAS-based multispectral sensors that are currently available. Tetracam offers a 6-band MCA6 multispectral imager covering the 450–1000 nm spectral region with a 38°-by-31° field-of-view. The MicaSense RedEdge-M sensor is a 5-band system (band centers at 475, 560, 668, 717, and 840 nm) that produces 12-bit RAW imagery with a 47.2° horizontal field-of-view (HFOV) (Barrows and Bulanon 2017; Potgeiter *et al.* 2017). Comparably, Sentera has developed the Multispectral Double 4K developed by Parrot (Paris, France) that is among these recently-released micro (remote) sensors. Primarily developed as an agricultural sensor, it has a five-lens architecture—four narrowband sensors (band centers at 550, with five channels (band centers at 446, 548, 650, 720, 840 nm) that utilizes a 12.3 MP backside illumination (BSI) complementary metal oxide semiconductor (CMOS) with a 60° HFOV. The Slanrange 3p imager is another four-channel system with configurable band centers between 410 and 950 nm.

Jarrod Edwards and John Anderson are with the USACE-ERDC Geospatial Research Laboratory, Corbin Field Station, 15319 Magnetic Lane, Woodford, Va. 22580 (Jarrod.D.Edwards@usace.army.mil). (John.Anderson@usace.army.mil).

William Shuart is with the Center for Environmental Studies, Virginia Commonwealth University, 1000 W. Cary St, Richmond Va. 22580 (wshuart@vcu.edu).

Jason Woolard is with the NOAA National Ocean Service, National Geodetic Survey, 15351 Office Dr, Woodford Va. 22580 (Jason.Woolard@noaa.gov).

Photogrammetric Engineering & Remote Sensing
Vol. 85, No. 3, March 2019, pp. 221–230.
0099-1112/18/221–230

© 2019 American Society for Photogrammetry
and Remote Sensing
doi: 10.14358/PERS.85.3.221

For this study, Parrot's Sequoia multispectral imager was used to collect orthorectified imagery in four multispectral bands. The Sequoia is largely a mapping system built around agricultural and vegetation remote sensing with bands centered at: 550 nm (green), 660 nm (red), 735 nm (red-edge), and 790 nm (near infrared). Each band is 40 nm full width at half maximum (FWHM) with the exception of the red-edge that is 10 nm FWHM. The sequoia also includes a companion red, green, blue (RGB) sensor for spatial reference. What makes the Sequoia unique is the addition of a flat, top-mounted irradiance sensor for recording down-welling solar flux (Handique *et al.* 2017; Jhan *et al.* 2017; Burud *et al.* 2017). In conjunction with an AIRINOV target panel these data are used within the PIX4D processing software (Lausanne, Switzerland) for post-processing reflectance calibration of the multispectral imagery.

While these miniature sensors are low-cost and convenient, it is important to remain objective and consider the need to fully understand how these data are acquired, corrected, and calibrated prior to product generation or analysis. Independent research is beginning to expose variability across systems. For example, Kelcev and Lucieer (2012) were able to correct sensor noise, vignetting, and lens distortion issues they discovered on the Tetracam mini-MCA. Von Bueren *et al.* (2015) found an issue with the Tetracam MCA6 in which the sensor was consistently overestimating the image green band when compared to spectra obtained from a ground radiometer. In addition, Jhan *et al.* (2017) found that parallax issues in both the Micasense red-edge and Sequoia multispectral imagers produced co-registration errors in the imagery. Many of these systems feature proprietary (cradle-to-grave) post-processing software to perform bundle adjustment and radiometric corrections essentially at the touch of a button. While convenient to the user, these correction features prevent the fundamental understanding of the data integrity and utility with (in the case of vegetation analysis) the need for multi-temporal acquisitions for comparative analysis where management decisions are critical.

This study attempted to investigate the spectral fidelity of drone-based multispectral imagery in three ways: First, we compared raw spectral intensities of radiometric ground targets between the band centers of the imaging sensor and corresponding wavelengths of our ground radiometer as a baseline analysis. Second, we compared the reflectance imagery generated by PIX4D (using the AIRINOV calibration standard)

to imagery corrected to reflectance using ground radiometer measurements of a Spectralon (National Institute of Standards and Technology (NIST)) standard and in-scene calibration targets by ELC. Finally, we compared two classifications using traditional error analysis between a ground truth data set and the AIRINOV-calibrated data.

Study Site

The site selected for this study was the National Oceanic and Atmospheric Administration (NOAA), National Ocean Service (NOS) geodetic surveying training facility at Corbin, Virginia (Figure 1). The site (Universal Transverse Mercator (UTM)), Zone 18N scene center 292060.83 m E, 4250963.83 m N) is a 120 hectare, semi-wooded and open site characterized by both deciduous and coniferous trees and rolling topography. While the data for the site represented a winter scene with broad areas of senescent vegetation (trees and grasses), there was enough cool weather grass and evergreen vegetation for a good spectral evaluation. The facility is a dedicated center for high accuracy geodetic mapping, unmanned aerial vehicle research, and testing and laser scanner evaluation. Geometric control targets used for this study were previously established during the summer of 2017 and surveyed using high precision GPS by rapid-static method resulting in sub-centimeter accuracies. All control data were processed using NOAA's National Geodetic Survey OnLine Positioning User Service (NGS OPUS) and the NovAtel Grafnet (Alberta, Canada) surveying package and placed in the WGS 84 horizontal datum in UTM Zone 18N coordinates using the NAVD 88 vertical datum.

Methodology

Radiometric Control

Radiometric control was established by using five ground targets (area = 0.36 m²) representing five spectral endmembers and were sized following Wang and Myint (2015). Reflectance spectra were acquired using two Spectra Vista Corporation (SVC, Poughkeepsie, New York) HR-1024i (350–2400 nm) ground radiometers between the mission times of 1135 and 1150 EST. The spectra obtained from both radiometers were averaged in order to reconcile the data between units and to reduce any anomalies present in either individual unit. Each target was placed and measured at or near an established control point during the airborne mission (Table 1). The targets consisted

of a series of gray paints (formulas by BEHR, Inc.) that are spectrally flat across the visible to the near infrared (VNIR) region of interest. The SVC was used to collect reflectance data in full sunlight at a distance of 1 m (nadir) above the ground target and (reflectance) calibration was performed by measuring a Spectralon (NIST) standard following procedures described by Satterwhite and Henley (1991). Target spectra are presented in Figure 2. Spectral resolution for the SVC instrument is ≤ 3.3 nm (FWHM) VNIR region and the nominal FOV of the instrument is 4 allowing a sample diameter of approximately 7 cm to be recorded. Five spectra were acquired per radiometer and averaged for each of the targets sampled to provide signatures used in the reflectance calibration of the Sequoia image data.

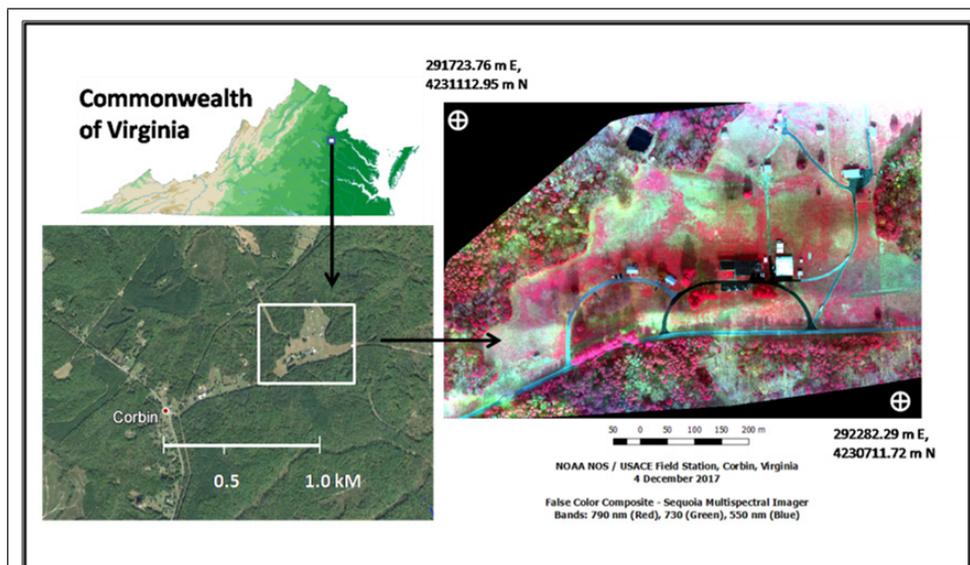


Figure 1. Study site location at NOAA/NOS Corbin, Virginia.

Airborne Imagery Acquisition and Geometric Rectification

High resolution remote sensing of the sites was accomplished using a fixed-wing Cumulus, semi-autonomous drone from Sky Watch (Støvring, Denmark) that served as the host platform for the Sequoia multispectral instrument (MSI) camera sensor. The drone was deployed late morning of 4 December 2017 between 1135 and 1150 EST. Sky conditions were generally clear and the sun angle was low (Start—Alt. 29.7°, EL. 172/End—Alt. 30.1°, EL. 177). A flying height of 90 m achieved a nominal 10 cm pixel ground sample distance (GSD) with a resulting frame imagery (side- and forward) overlap of 70%. The resulting data (300 imagery frames) were photogrammetrically processed using ground control to

produce orthorectified spectral imagery and a composite possessing four spectral bands: 550 nm, 660 nm, 735 nm, 790 nm within the PIX4D processing software package.

Imagery Reflectance Calibration

Reflectance calibration of the PIX4D-processed data involved the use of an AIRINOV grey calibration standard (Figure 3) as part of the recommended Sequoia image correction procedure within PIX4D. These corrections were applied to the raw 16-bit data resulting in four separate reflectance-calibrated GEOTIFFs. Procedurally, this involved the pre- and post-flight collection of an image of the target standard and is assumed to transform the raw data from the sensor to percent reflectance in conjunction with data from the down-welling sensor located on top of the imager that provides solar irradiance (Parrot, 2017). It is not clear how this is performed, but two procedures are possible: 1) a simple ratio of the upwelling to the down-welling measurements or 2) computing the reflectivity for the band centers and then squaring those terms to derive the reflectance (McCluney, 1994).

To create a reflectance imagery data set based on ground truth, radiometric correction models were developed using the raw 16-bit Sequoia sensor data and the SVC ground radiometry to generate modified Empirical Line Calibration (ELC) corrections for UAS camera data following Wang *et al.* (2015) and Smith *et al.* (1999). Ground radiometric measurements were collected concurrently within the mission time of the flight in order to acquire the target and feature spectra being imaged. ELC was performed within the open source Opticks spectral processing package (United States Air Force and Ball Aerospace). This correction provided four separate reflectance-calibrated GEOTIFFs, each based on the ground spectral signatures at their respective band center.

Table 1. Calibration target list and UTM geographic position.

Target & Avg % Reflectance	Description	Easting (m)	Northing (m)
Gray 1—56.0	Bright gray endmember	292 220.73	4 230 925.19
Gray 2—25.20	Intermediate gray endmember	291 943.00	4 230 888.68
Gray 3—16.00	Intermediate gray endmember	292 121.32	4 231 022.61
Gray 4—7.00	Dark gray endmember	292 177.72	4 230 833.62
Spectralon—99.9	NIST bright white endmember	292 480.72	4 230 875.44

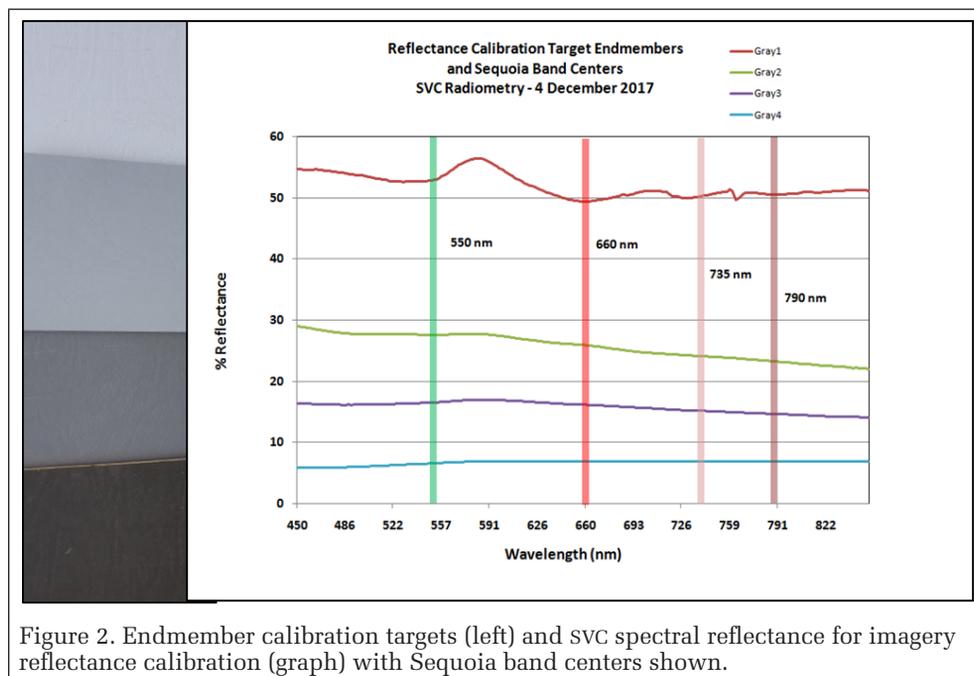


Figure 2. Endmember calibration targets (left) and SVC spectral radiometry for imagery reflectance calibration (graph) with Sequoia band centers shown.

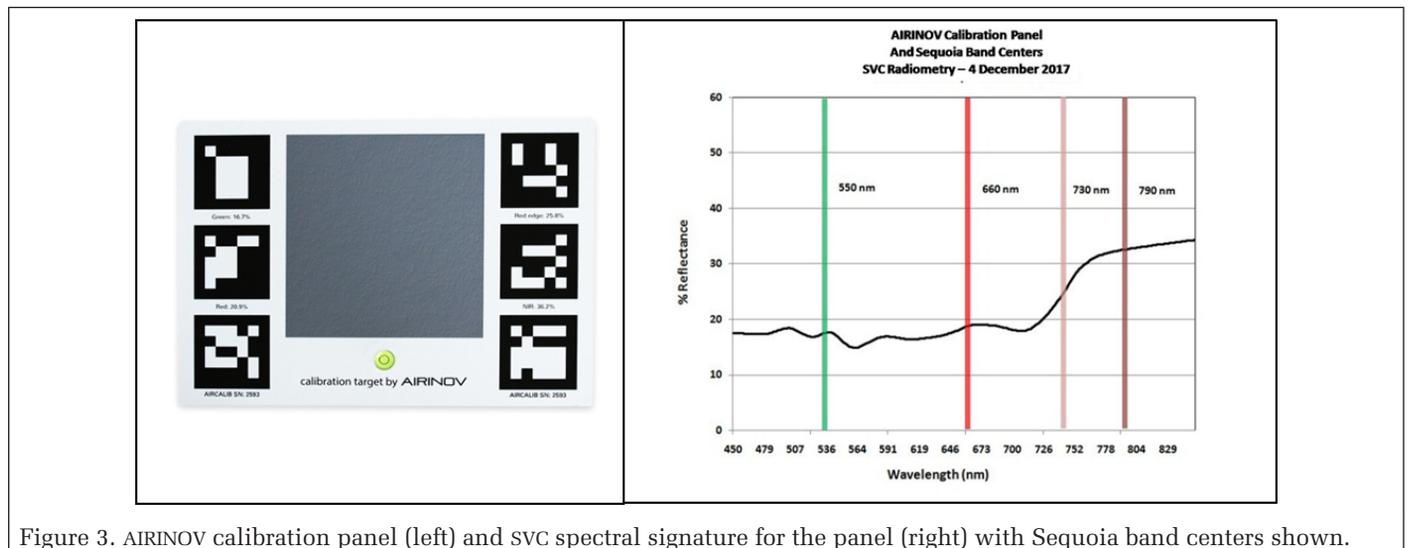


Figure 3. AIRINOV calibration panel (left) and SVC spectral signature for the panel (right) with Sequoia band centers shown.

Spectral Analysis and Classification of Features

Sequoia imagery, geospatial layers (e.g., training sample shapefiles) and spectral reflectance data were entered into the Quantum Geographic Information System (QGIS) (ver. Nodebo 2.16) and saved as project files. Pixels from both the AIRINOV and SVC calibrations were queried to obtain reflectance values for each imagery data set and statistically compared to the ground reflectance and each other. The signature processing, training data extraction, training sample analysis resulting classifications, and error analyses were executed with QGIS using the Semi-Automatic Classification Plug-In (SCP) developed by Luca Congedo (2016). SCP allows ingestion, exploitation, production, and analysis of multiband data and spectral signatures along with several robust classifiers including the Spectral Angle Mapper (SAM) algorithm used with our data sets. SAM is an automated method for comparing image spectra to individual or library spectral data and assumes that the data have been reduced to *apparent* reflectance (e.g., atmospherically corrected). The SAM algorithm determines class spectral similarity by calculating the angle between them, treating them as vectors in a space dimensionally equal to the number of bands in the data set (Rashmi *et al.* 2014; Kruse *et al.* 1993a). The inputs to the SAM classifier were training areas developed from the imagery that included the calibration targets and ground features: vegetation (coniferous and deciduous trees and grasses), water, pavements, and structures that were saved as shapefiles. Twelve training samples were extracted from each image (AIRINOV and SVC) that included both targets and broad ground features using the region of interest (ROI) tool in SCP. The area extracted for the endmembers and class ROIs was centered on each feature and covered 0.35 m² in area, encompassing the target and maintaining high pixel purity (Figure 4). Polygonal ROIs for the feature classes varied in size and dimension capturing as much of an area as possible to spectrally and statistically characterize it. This assured the sample area was completely represented in the training statistic. Prior to running the SAM classifier, spectral signatures were generated for each training sample and analyzed for statistical separation using the spectral angle method. The spectral angle method returns results representing degrees of separation for training samples where scores close to 0 are identical (same feature) and scores close to 180 represent sample features that are unique. This analysis is part of the classification procedure and provides confidence in target/signature selection.

Error and Statistical Analysis

While the range of camera silicon detectors are nonlinear in spectral response (300 nm to 1100 nm), the range of our investigation is limited to band centers representing a nearly

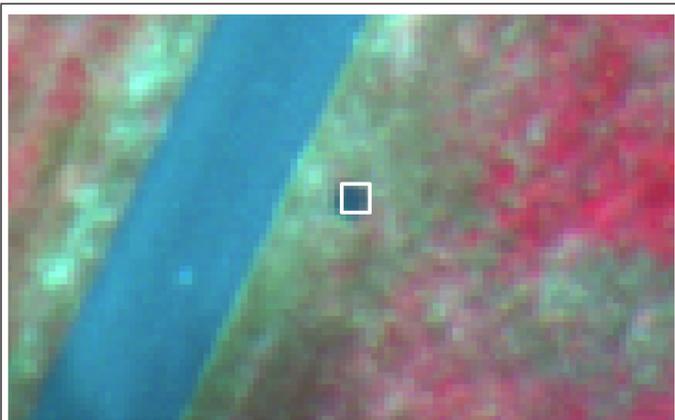


Figure 4. ROI polygon for endmember (Gray Target 4).

linear part of the response curve that includes a plateau from 790 to 800 nm (Sze, 1985). Comparisons between the imagery data reflectance and ground spectral reflectance data were performed using simple linear regression analysis and model relationships are reported as r^2 . For the observed and measured reflectance extracted from target pixels (between the AIRINOV and SVC-calibrated data), the root mean square error (standard deviation of the residuals) was computed for differences in percent reflectance between models. Map classification accuracies for targets and features were computed between the AIRINOV and SVC (truth) class maps by generating errors of omission and commission (confusion matrix) following procedures described by Congalton (1991) and Campbell (2011). The error matrix was developed within the SCP image processor using the SVC map as the reference (ground truth) and the AIRINOV map as the map to be assessed.

Results

Sequoia Image Reflectance Calibrations and Multispectral Orthomosaics

As described, two data sets were produced from the Sequoia multispectral imagery representing the AIRINOV calibration

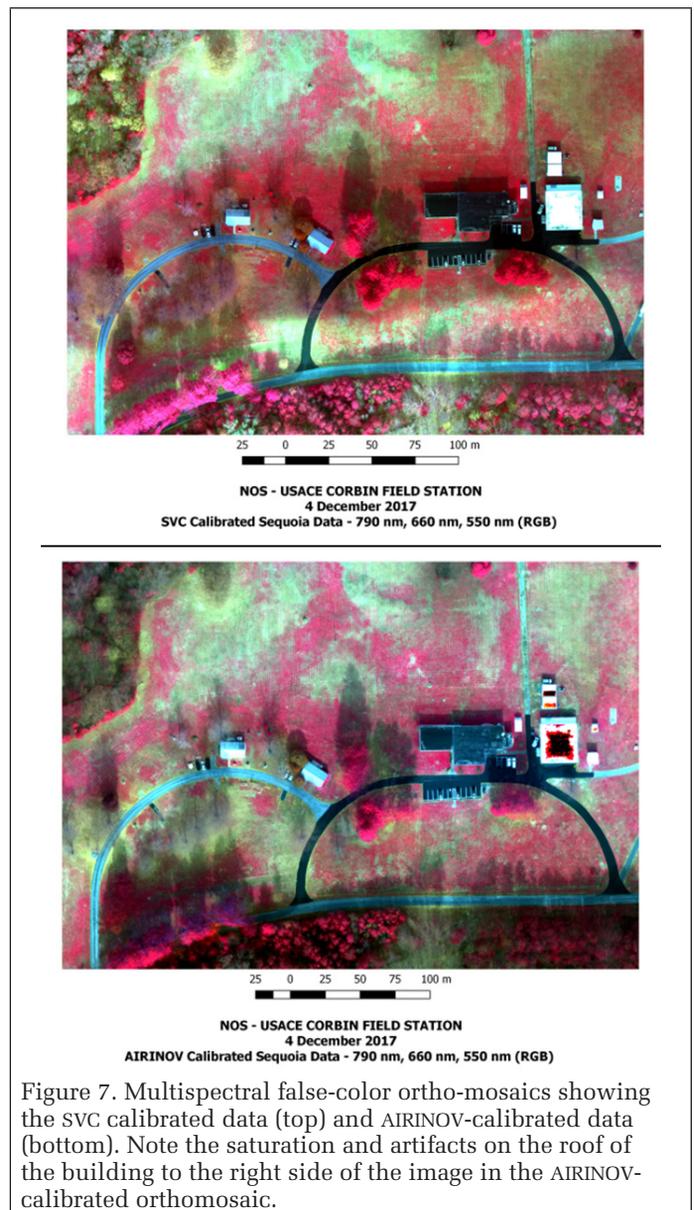


Figure 7. Multispectral false-color orthomosaics showing the SVC calibrated data (top) and AIRINOV-calibrated data (bottom). Note the saturation and artifacts on the roof of the building to the right side of the image in the AIRINOV-calibrated orthomosaic.

and the SVC radiometric data calibration by ELC acquired during the mission. Spectral models are shown in Figure 5 relating the ground reflectance between the SVC and Sequoia (raw 16 bit) data. These data are the result of comparing the ground radiometry of the targets acquired by the SVC to the *un-reflectance calibrated* Sequoia orthorectified imagery across all four multispectral bands.

Figure 6 presents the models relating the AIRINOV-calibrated Sequoia data (post-processed using the PIX4D correction procedure) and the SVC target ground data for each band. Figure 7 shows the resulting (orthomosaic) false color composites resulting from the SVC and AIRINOV corrections. Figure 8 shows the relationships between the SVC calibrated Sequoia imagery as compared to the ground spectral reflectance. For each of the models (both AIRINOV and SVC image data) the RMSE was computed to derive the (%) reflectance and these results are given in Table 2.

Signature Separation Analysis of Training Samples

Figure 9 shows band-versus-band (separation) positions in spectral space. Resulting signature separation (SAM) scores for terrain features represented values ranging from 4.63 (senescent grass and deciduous trees) and 10.40 (coniferous trees and green grasses) to 159.14 (pavement dark and bright urban features).

Once the separation analysis was concluded, the training samples were submitted to the SAM classifier. SAM is discrete and based on the measurement of the similarity between two spectra for a given feature across all bands. This similarity can be obtained by considering each spectrum as a vector in q -dimensional space, where q is the number of bands as a vector in q -dimensional space, where q is the number of bands.

Classification Maps and the SVC–AIRINOV Error Matrix

The resulting classification maps for the twelve classes (plus an unclassified layer) are shown in Figures 10 and 11. Both classifications show good representations of the features based upon the sample training statistics, however there are some differences that appear influenced by the calibration procedures. Errors of omission and commission were performed using the SVC-calibrated classification map as a reference (truth) data set in an effort to analyze the AIRINOV class map result. Table 3 presents these results. When compared to the SVC-calibrated data, the PIX4D

Table 2. Root mean square errors for Sequoia imagery calibration models showing AIRINOV and SVC using measured ground data results (reference Figures 6 and 8).

Sequoia Imagery Band (nm)	AIRINOV		SVC	
	R ²	RMSE (%Ref)	R ²	RMSE (%Ref)
550 nm	0.43	8	0.99	1
660 nm	0.30	12	0.88	10
735 nm	0.59	8	0.97	5
790 nm	0.96	6	0.98	4

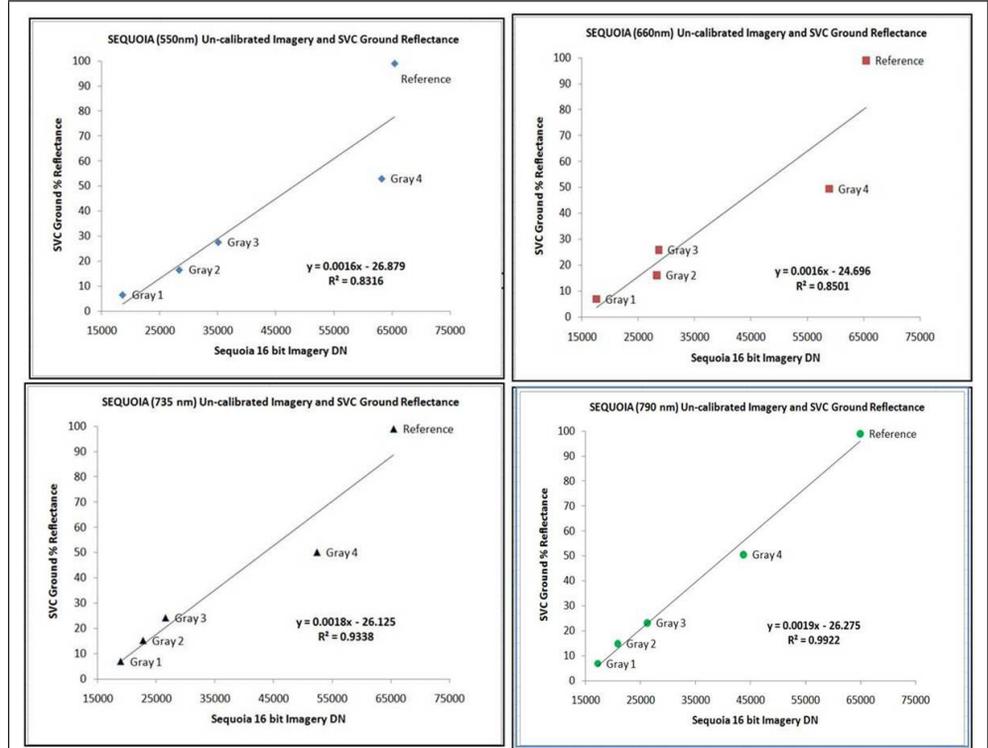


Figure 5. Reflectance models comparing the un-calibrated Sequoia multispectral imagery bands (in digital numbers) and the SVC ground radiometric data for the targets.

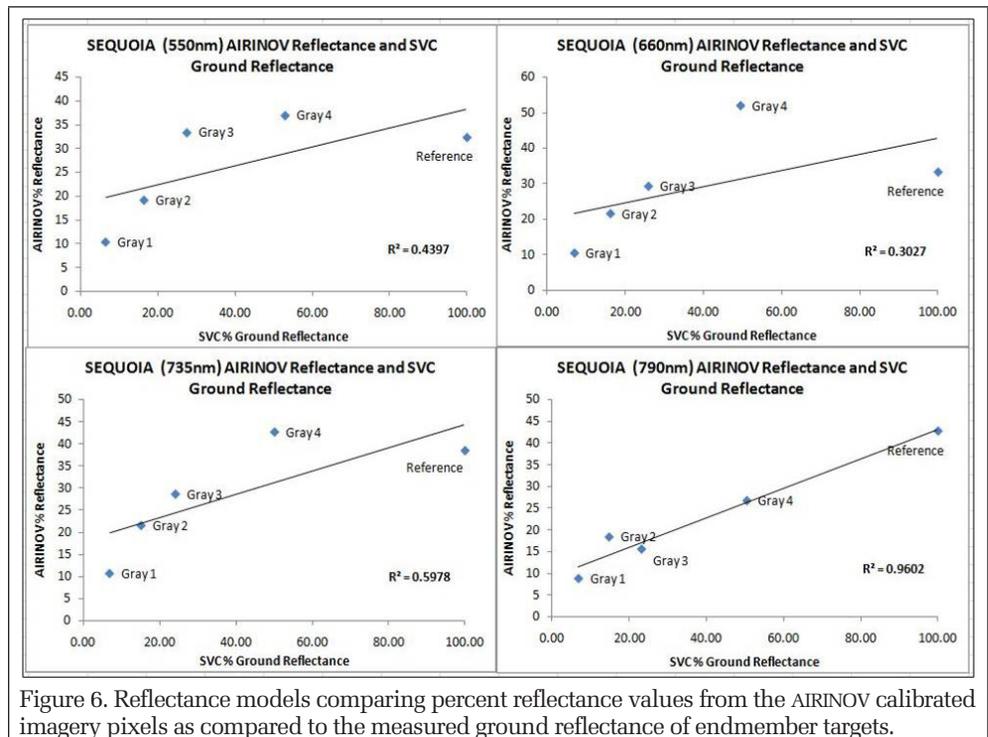


Figure 6. Reflectance models comparing percent reflectance values from the AIRINOV calibrated imagery pixels as compared to the measured ground reflectance of endmember targets.

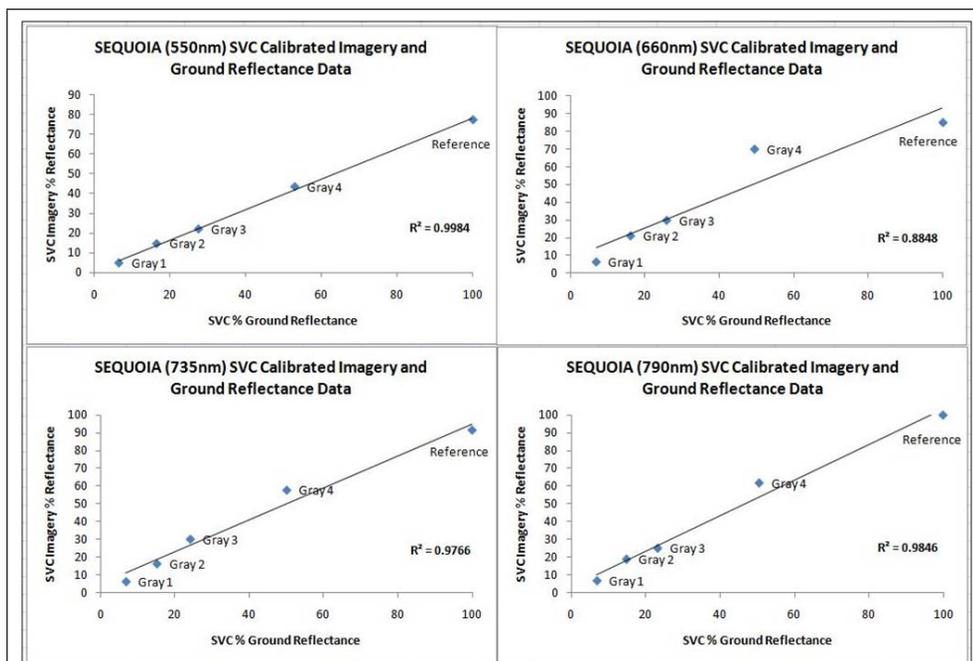


Figure 8. Reflectance relationships between SVC-calibrated imagery pixel reflectance and ground reflectance of endmember targets.

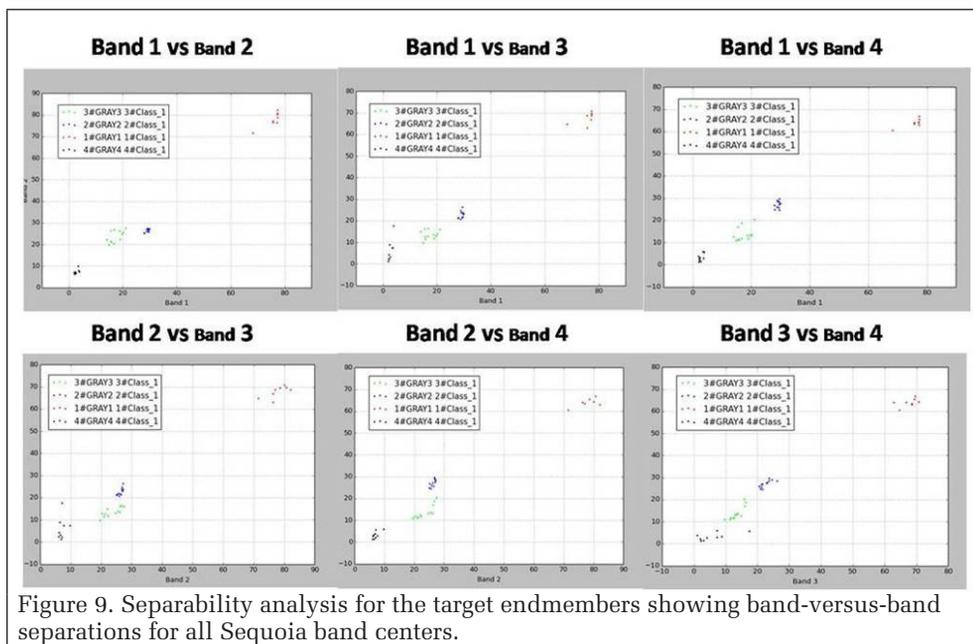


Figure 9. Separability analysis for the target endmembers showing band-versus-band separations for all Sequoia band centers.

AIRINOV—calibration achieved only 24% of the classes based upon the same training shapefiles with a considerable number of unclassified pixels. A difference map (Figure 12) was also generated to show the broad areas of agreement and disagreement between the class maps.

Discussion

Our exploration of the spectral discretion in a small-format multispectral imaging system using two different calibration methods provided a unique opportunity to expose the need for standardization in (radiometric) processing. As stated, these systems are becoming popular in the remote sensing community for surveying and mapping and their radiometric attributes are under-studied. While not a particularly ideal time of year to evaluate a system designed around agricultural uses, the Sequoia MSI demonstrated good feature mapping capabilities that may be improved with a modified (and verifiable) reflectance calibration workflow.

In our study using this small-format system, we attempted to draw conclusions related to the comparison of a ground-truth data set based on *in situ* spectral measurement and the advantageous collection of airborne radiometric data combined with a calibration standard. Our initial analysis showed very good agreement between the raw, un-calibrated Sequoia data and ground radiometer measurements among the ground targets, producing models and r^2 values that were high in each of the four bands featured by the system. However, the difference between the AIRINOV-calibrated imagery data using the PIX4D software correction

Table 3A. Error Matrix for PIX4D–AIRINOV.

	Un-classified	GRAY Target 1	GRAY Target 2	GRAY Target 3	GRAY Target 4	Senesced Grass	Healthy Grass	Deciduous Trees	Coniferous Trees	Urban	Pavement Light	Pavement Dark	Open Water	Total
Unclassified	15 483	4884	17	99	2228	2669	2125	1957	1805	640	90	3 322 563	3 354 560	6 709 120
GRAY Target 1	4715	5267	32 792	3222	5662	1384	8806	1551	34 864	5532	2203	2827	108 825	217 650
GRAY Target 2	13 132	13 897	4709	392	1523	559	881	367	12 599	2583	873	419	51 934	103 868
GRAY Target 3	4428	2428	10 493	1282	11 323	3306	13 066	2712	5289	3199	2073	615	60 214	120 428
GRAY Target 4	9723	13 523	12 389	2353	106 368	51 018	67 752	28 982	4929	6040	14542	3044	320 663	641 326
Senesced Grass	3946	6254	4921	2590	980 062	255 299	806 140	270 270	773	5072	8730	2447	2 346 504	4 693 008
Healthy Grass	593	1983	527	620	357 181	1 092 279	160 808	1 817 347	320	2894	67334	3907	3 505 793	7 011 586
Deciduous Trees	3612	10 633	3312	3169	1 556 805	1 298 372	1 136 653	1 634 694	1325	8646	92 527	4817	5 754 565	11 509 130
Coniferous Trees	201	374	262	140	18 550	460 391	16 938	727 956	244	1179	20872	2701	1 249 808	2 499 616
Urban	8895	12 920	16 958	899	981	771	989	446	32 714	9336	63 355	11 955	160 219	320 438
Pavement Light	1912	3447	36 362	5563	2112	1356	3482	693	47 713	7324	19 076	27 691	156 731	313 462
Pavement Dark	939	1247	10 408	2379	2823	5064	3648	2349	6169	1749	21 551	7962	66 288	13 2576
Open Water	631	570	925	561	264	279	227	169	1632	738	45 326	42 574	93 896	187 792
Total	68 210	77 427	134 075	23 269	3 045 882	3 172 747	2 221 515	4 489 493	150 376	54 932	358 552	3 433 522	17 230 000	34 460 000

Overall accuracy [%] = 23.636.

Table 3B. User and Producers Error for PIX4D-AIRINOV.

	Producer Accuracy (%)	User Accuracy (%)	Kappa hat/Statistic
Gray Target 1	6.912 476 176 51	4.332 644 153 46	0.039 524 116 519 4
Gray Target 2	17.948 519 250 4	26.758 963 299 6	0.264 283 520 409
Gray Target 3	7.826 216 669 77	17.426 179 958 1	0.167 785 937 689
Gray Target 4	10.112 166 401 7	0.733 792 174 339	0.005 995 525 334 74
Senesced Grass	32.176 624 045 2	41.766 900 887 4	0.292 619 888 167
Grass	34.426 917 746 7	31.156 403 130 5	0.156 182 808 93
Deciduous Trees	51.165 668 474	19.752 196 734 2	0.078 741 358 458 9
Coniferous Trees	16.214 659 428 1	58.245 426 497 5	0.435 319 723 581
Urban	21.754 801 298 1	20.418 302 448 5	0.197 176 326 123
Pavement Light	13.332 847 884 7	4.672 974 714 64	0.043 680 848 502 7
Pavement Dark	6.010 564 715 86	32.511 163 408 2	0.310 768 907 045
Open water	1.239 951 280 35	45.341 654 596 6	0.317 388 618 094

and the traditional (Empirical Line) method based upon the SVC ground reflectance spectral catalogs was stark. Regression analysis revealed both visible bands (550 nm and 660 nm) r^2 yields of less than 0.44, with the 790 nm (NIR) band possessing the best fit at 0.96 for the AIRINOV-calibrated data. With the exception of the NIR (790 nm) band, the AIRINOV-calibrated imagery showed relationships roughly half as significant as those resulting from the SVC-calibrated imagery band for band (Table 2). Regression analysis for the SVC-calibrated imagery returned strong correlations with r^2 yields of no less than 0.88.

The computed RMSE for both imagery data sets supported the regression results, with the AIRINOV producing significantly larger reflectance deviations than the SVC-calibrated data (Table 2). The most pronounced deviations in RSME, regardless of calibration method, were in the 660 nm (red) band. In both the AIRINOV and SVC-calibrated imagery the reflectance RMSE results were 12% and 10%, respectively for the 660 nm band with some saturation and artifacts introduced into the AIRINOV imagery. This result has the potential of confounding soil, vegetation and relevant feature reflectance in a critical spectral band for characterization, an issue also described by Von Bueren *et al.* (2015) for comparable MCA6 visible data. Original research by Tagle (2017) has reported sources of error in irradiance compensation due to location and design of down-welling sensors used in some UAV multispectral camera systems. In addition, the use of a diffuser, instead of a hemispheric view can result in differences up to 44% when compared to ground spectral measurements profoundly affecting the post calibration of data (Hakala *et al.* 2013). This may help explain why the RMSE is also high (10%) for the SVC calibrated 660 nm image but exhibits neither saturation, nor artifacts. After further inspection of this spectral band, there was no apparent registration or appreciable vignetting (as described by Yu (2004) that could contribute to such a result. Figure 13 presents a comparison of the Sequoia imagery bands including the: 16 bit uncalibrated data, AIRINOV calibrated imagery, and SVC (ELC) calibrated imagery. As discussed, post-calibration saturation and artifacts introduced into the AIRINOV data served to contribute to a large number of unclassified pixels, mostly from bright-toned soils and building materials (see Figure 11).

Upon further examination of the AIRINOV-calibrated reflectance imagery, regression plots (see Figure 6)

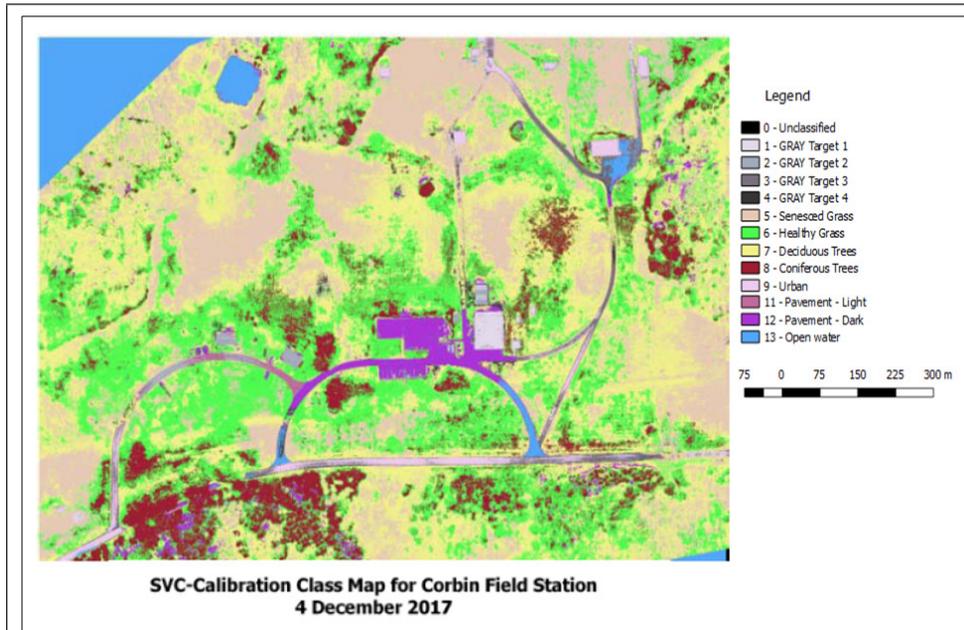


Figure 10. SVC class map for twelve classes based upon the SVC-reflectance calibrated orthorectified imagery.

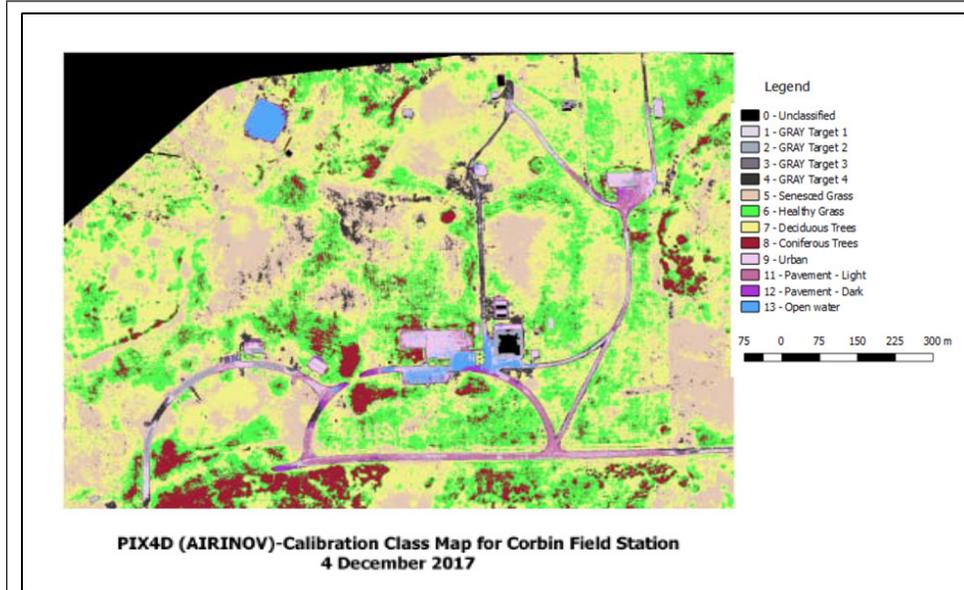


Figure 11. AIRINOV class map for twelve classes based upon the PIX4D-reflectance calibrated orthorectified imagery.

clearly show a drop in intensity values for the white (NIST) reference standard that should be characterized as roughly 100% reflectance. In fact, the intensity values for the white standard were lower than the brightest gray target in three of the four bands (550, 660, 735 nm) suggesting inadequate calibration resulting in a reflectance threshold below the brightest endmembers. Such an issue could potentially produce incorrect spectral characterization for bright tone soils and plant vigor in far-red and infrared bands. Removal of the white

reference data point in the AIRINOV regressions drastically improved the linear models as presented in Figure 14. This improvement in linear fit was most pronounced in the 660 nm band where r^2 changed from the worst fit among the AIRINOV bands ($r^2 = 0.30$) to the best fit ($r^2 = 0.99$). This was not the case with the un-calibrated, 16-bit DN data from the Sequoia, in which all bands displayed more of the expected brightness progression among targets (reference Figure 5).

These results were unexpected for a system that applies a combination of a down-welling sensor and calibration standard for the derivation of reflectance. In the use of many small format systems using commercial cameras, typical adjustments for f-stop and integration time are made prior to the mission, but changing illumination conditions can make these data cumbersome to calibrate radiometrically. It is certainly an advantage of this (and similar) multispectral camera system(s) that offer the measurement of true, real-time radiometric data applied in post-processing for reflectance correction. This is particularly important in systems possessing a small number of band-sets that are "tailored" to specific applications (i.e., agricultural).

In order to evaluate our findings in a real-world application, we generated a classification map using ground-truth radiometric measurements to supervise classification in imagery using the AIRINOV-calibrated dataset. Fundamentally, issues involving these kinds of differences in feature reflectance are compounded during classification and reduce the accuracy of analytical products and temporal comparability. As previously observed, it is unknown which reflectance correction method is applied in conjunction with the down-welling data and AIRINOV standard for the reflectance calibration within PIX4D, but in our experiment we purposefully generated class maps based upon features for which ground spectral measurements were collected. The

overall accuracy of the AIRINOV map as compared to the SVC (truth) map was 24%, indicating much room for improvement for data collected at the same time on the same day under ideal conditions. The SAM classification generated for each calibration procedure produced varied results, concluding in fair to excellent separation of the calibration targets (reference Figure 10). However, the feature classification (as compared to the ground truth) was poor for the AIRINOV reflectance calibrated imagery.

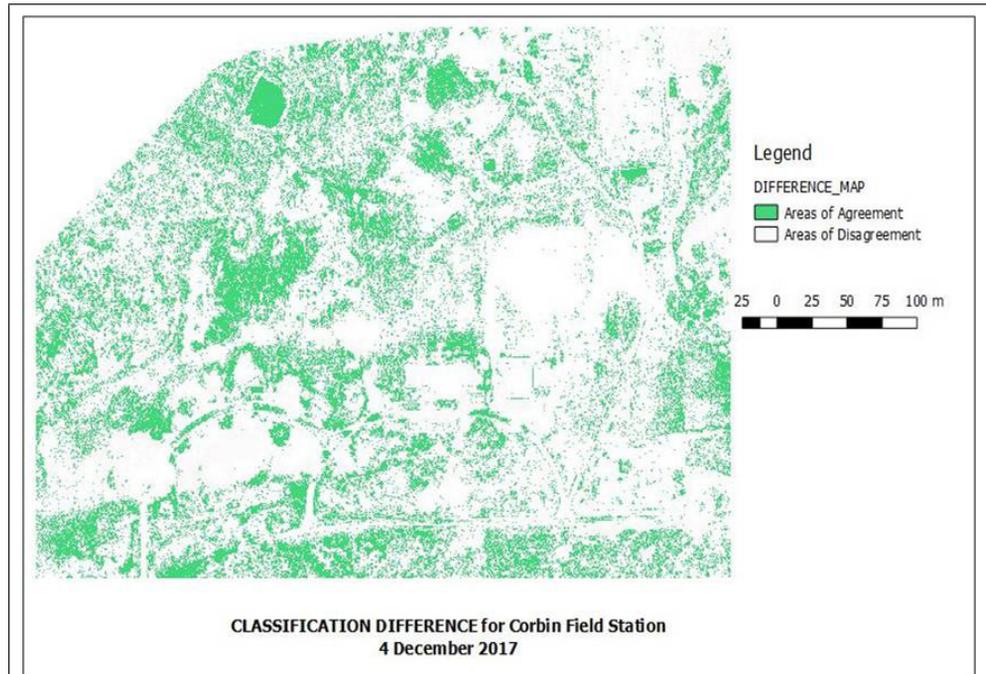


Figure 12. Difference Map showing areas of agreement (and disagreement) between the AIRINOV map and the SVC (truth) maps.

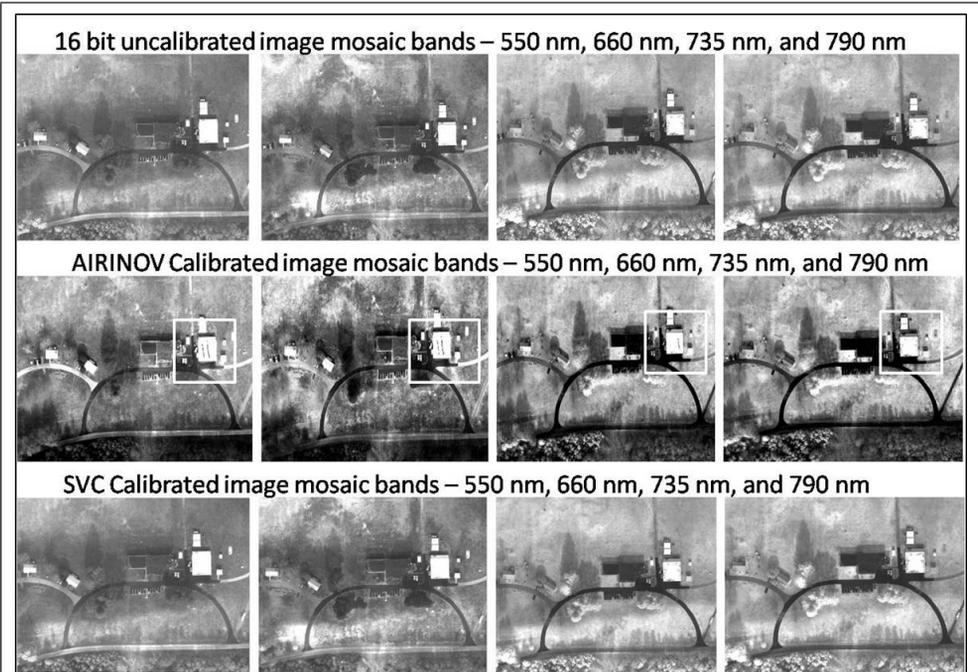


Figure 13. Imagery band subsets from the Sequoia multispectral imagery mosaic. Note the area (box) in the AIRINOV imagery depicting an extreme example of saturation and artifacts introduced after reflectance calibration. These areas extend to bright toned soils and some vegetated areas as well.

Conclusion

The objective of this study was to compare spectral data from a small-format multispectral remote sensing imager by testing two different calibration methods to better understand the spectral fidelity of the system. Integral to any spectral analysis or classification is the integrity of the data and how it accurately relates to ground features in reflectance space at the time of acquisition. This is fundamental to accurate classification and subsequent map production, especially if multi-temporal data will be necessary as with seasonal acquisitions. We used two reflectance calibration workflows to accomplish the evaluation of data and products with a popular, commercially available multispectral sensor. While the quality of the raw imagery was excellent, the post-processing and reflectance calibration method using the manufacturer-suggested AIRINOV calibration did not compare favorably to the ELC-reflectance calibration based on spectral endmembers for radiometric adjustment. This finding suggests that the imager itself produces accurate radiometric

data, however the manufacturer-suggested reflectance calibration method (AIRINOV calibration) is likely the source of many of the reflectance discrepancies we observed in this study. Ultimately, the quality of data and products produced from UAV spectral sensors depends on standardized acquisition and correction processes to accurately reflect radiometric performance so comparable data sets can be realized.

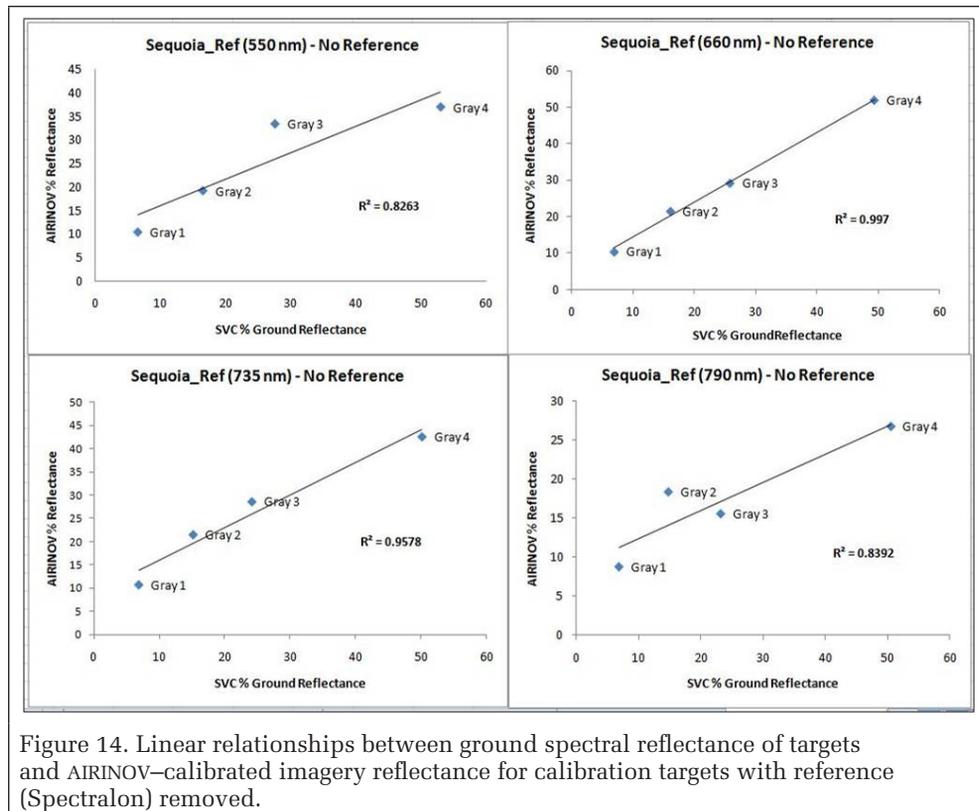


Figure 14. Linear relationships between ground spectral reflectance of targets and AIRINOV-calibrated imagery reflectance for calibration targets with reference (Spectralon) removed.

Since the Sequoia is built around vegetation analysis, and given the late time of year, vegetation feature differences were only prominent between green, cold weather grasses (Healthy Grass) and dormant grasses (Senesced Grass), and deciduous trees (stems) and coniferous stands of pine. Separation of some map features (e.g., urban, pavements, and open water) were fair to excellent; however, because of the limited number of spectral bands, reflectance calibration procedure, viewing geometry, target anisotropy, and other ancillary factors related to the sensor itself served to foster errors into the resulting classification. For example, both class maps possessed commission errors, particularly with attention to the Open Water class. Of interest is that in the SVC-based map, the Urban category omits pixels associated with buildings having asphalt roofs and categorizes those as Pavement Dark (a dark asphalt feature). The AIRINOV map identifies these features as Urban with the addition of some Pavement Dark and Open Water pixels. Other areas containing sizeable commission errors include the Healthy Grass and Coniferous Trees classes and the Senesced Grass and Deciduous Trees classes. The classes representing deciduous trees, dark pavement, and open water also possessed higher Kappa scores in the error analysis. We speculate that a contributing factor to the classification differences originated with the standards and methods used in the reflectance calibration. Furthermore, the AIRINOV target is a uniformly gray standard with a signature close to cellulose, low in the visible bands, but high in the red-edge and near IR (see Figure 3). The changes in amplitude across the available wavelengths may introduce inconsistency and variability in the calibration, especially when combined with the other factors mentioned that serve to affect feature separability. In this experiment, the ELC method, when used with a concurrent spectral survey of endmembers and a standard uniformly bright target (e.g., NIST standard) was more effective in the reflectance post processing of the data. This was demonstrated in imagery free of artifacts and extreme saturation and a resulting classification where all classes were populated.

References

- Barrows, C. and D. M. Bulanon. 2017. Development of a low-cost multispectral camera for aerial crop monitoring. *Journal of Unmanned Vehicle Systems* 5 (4):192–200.
- Burkart, A., H. Aasen, L. Alonso, G. Menz, G. Bareth, U. Rascher. 2015. Angular dependency of hyperspectral measurements over wheat characterized by a novel UAV based goniometer. *Remote Sensing* 7:725–746.
- Burud, I., G. Lange, M. Lillemo, E. Bleken, L. Grimstad and P. Johan. 2017. Exploring robots and UAVs as phenotyping tools in plant breeding *IFAC PapersOnLine* 50 (1):11479–11484.
- Campbell, J. B. and R. H. Wynne. 2011. *Introduction to Remote Sensing*. 408–418. New York, N.Y.: Guilford Press.
- Chander, G. and B. L. Markham. 2003. Revised Landsat-5 TM radiometric calibration procedures and postcalibration dynamic ranges. *IEEE Transactions on Geoscience and Remote Sensing* 41 (11):2674–2677.
- Chander, G., B. L. Markham and J. A. Barsi. 2007. Revised Landsat-5 thematic mapper radiometric calibration. *IEEE Geoscience and Remote Sensing Letters* 4 (3):490–494.
- Chander, G., B. L. Markham and D. L. Helder. 2009. Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote Sensing of Environment* 113 (5):893–903.
- Cocks, T., R. Jenssen, A. Stewart, I. Wilson and T. Shields. 1998. The HyMap airborne hyperspectral sensor: The system, calibration and performance. pp. 37–42 in *Proceedings at 1st EARSel Workshop on Imaging Spectroscopy*, held in Zurich, Switzerland, 6–8 October 1998.

- Colomina, I. and P. Molina. 2014. Unmanned aerial systems for photogrammetry and remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing* 92:79–97.
- Congalton R. G. 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment* 37 (1):35–46.
- Cowley, D. C., C. Moriarty, G. Geddes, G. L. Brown, T. Wade and C. J. Nichol. 2018. UAVs in context: Archaeological airborne recording in a national body of survey and record. *Drones* 2(2).
- Eismann, M. 2012. *Hyperspectral Remote Sensing*, 746. Bellingham, Wash.: Society of Photo-Optical Instrumentation Engineers.
- Elias, B. 2012. Pilotless drones: Background and considerations for congress regarding unmanned aircraft operations in the National Airspace System. Washington, D.C.: CRS Report for Congress, R42718. <<https://fas.org/spp/crs/natsec/R42718.pdf>> Accessed 10 September 2012.
- Gao, B., M. Montes, C. Davis and A. F. H. Goetz. 2009. Atmospheric correction algorithms for hyperspectral remote sensing data of land and ocean. *Remote Sensing of Environment* 13:S17–S24.
- Haboudane, D., J. R. Miller, E. Pattey, P. J. Zarco-Tejada and I. B. Strachan. 2004. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sensing of Environment* 90:337–352.
- Hakala, T., E. Honkavaara, H. Saari, J. Mäkynen, J. Kaivosoja, L. Pesonen and I. Pölönen. 2013. Spectral imaging from UAVs under varying illumination conditions. *ISPRS International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XL-1/W2:189–194.
- Handique, B. K., A. Q. Khan, C. Goswami, M. Prashnani, C. Gupta and P. L. N Raju. 2017. Crop discrimination using multispectral sensor onboard unmanned aerial vehicle. *Proceedings of the National Academy of Sciences, India—Section A: Physical Sciences* 87 (4):713–719.
- Jhan, J. P., J. Y. Rau, N. Haala and M. Cramer. 2017. Investigation of parallax issues for multi-lens multispectral camera band co-registration. pp. 157–163 in *Proceedings of the International Conference on Unmanned Aerial Vehicles in Geomatics*, held in Bonn, Germany, 4–7 September 2017.
- Justice, A. B., J. Morisette, P. Lewis, J. Privette and F. Baret. 2000. Developments in the validation of satellite sensor products for the study of the land surface. *International Journal of Remote Sensing* 21:17.
- Kelcey, J. and A. Lucieer. 2012. Sensor correction of a 6-band multispectral imaging sensor for UAV remote sensing. *Remote Sensing* 4:1462–1493.
- Kruse, F. A., A. B. Lefkoff, J. W. Boardman, K. B. Heidebrecht, A. T. Shapiro, J. P. Barloon and A. F. H. Goetz. 1993. The spectral image processing system (SIPS)—Interactive visualization and analysis of imaging spectrometer data. *Remote Sensing of Environment* 44 (2–3):145–163.
- McCluney, W. R. 2014. *Introduction to Radiometry and Photometry*, 2nd ed. 402. Norwood, Mass.: Artech House.
- Mishra, N., D. Helder, J. Barsi and B. Markham. 2016. Continuous calibration improvement in solar reflective bands: Landsat 5 through Landsat 8. *Remote Sensing of Environment* 185:7–15.
- Mumby, P. J., E. P. Green, C. D. Clark and A. J. Edwards. 1998. Digital analysis of multispectral airborne imagery of coral reefs. *Coral Reefs*, 17:59–69.
- Potgieter, A. B., B. George-Jaeggli, S. C. Chapman, K. Laws, L. A. Suárez Cadavid, J. Wixted, J. Watson, M. Eldridge, D. R. Jordan and G. L. Hammer. 2017. Multi-spectral imaging from an unmanned aerial vehicle enables the assessment of seasonal leaf area dynamics of sorghum breeding lines. *Frontiers in Plant Science*, 8:1532.
- Price, J. C. 1987. Radiometric calibration of satellite sensors in the visible and near infrared: History and outlook. *Remote Sensing of Environment* 22 (1):3–9.
- Rashmi, S., A. Swapna, S. Venkat and S. Ravikiran. 2014. Spectral angle mapper algorithm for remote sensing image classification. *International Journal of Innovative Science, Engineering and Technology* 1 (4):201–205.
- Richter, R. and D. Schläpfer. 2002. Geo-atmospheric processing of airborne imaging spectrometry data. Part 2: Atmospheric/topographic correction. *International Journal of Remote Sensing* 23(13).
- Satterwhite, M. B. and J. P. Henley. 1990. Hyperspectral signatures (400 to 2500 nm) of vegetation, minerals, soils, rocks, and cultural features: Laboratory and field measurements. Fort Belvoir, Va: US Army Corps of Engineers, ETL-0573.
- Schaepman-Strub, G., M. E. Schaepman, T. H. Painter, S. Dangel and J. V. Martonchik. 2006. Reflectance quantities in optical remote sensing—definitions and case studies. *Remote Sensing of Environment* 103 (1):27–42.
- Smith, G. M. and E. J. Milton. 1999. The use of the empirical line method to calibrate remotely sensed data to reflectance. *International Journal of Remote Sensing* 20 (13):2653–2662.
- Stark, B., T. Zhao and Y. Chen. 2016. An analysis of the effect of the bidirectional reflectance distribution function on remote sensing imagery accuracy from small unmanned aircraft systems. pp. 1342–1350.C in *Proceedings of the International Conference on Unmanned Aircraft Systems*, held in Arlington, Va, 7–10 June 2016.
- Sze, S. M. 1985. *Semiconductor Devices Physics and Technology*. 257. New York, N.Y.: Wiley.
- Tagle, X. 2017. Study of radiometric variations in unmanned aerial vehicle remote sensing imagery for vegetation mapping. MS thesis, Lund University, Lund, Netherlands, 64.
- Thome, K., D. Gellman, R. Parada, S. Biggar, P. Slater and M. S. Moran. 1993. In-flight radiometric calibration of Landsat-5 Thematic Mapper from 1984 to the present. In *Proceedings of SPIE 1938, Recent Advances in Sensors, Radiometric Calibration, and Processing of Remotely Sensed Data*, 15 November 1993.
- Von Bueren, S. K., A. Burkart, A. Hueni, U. Rascher, M. P. Tuohy and I. J. Yule. 2015. Deploying four optical UAV-based sensors over grassland: challenges and limitations. *Biogeosciences* 12:163–175.
- Wang, C. and S. Myint. 2015. A simplified empirical line method of radiometric calibration for small unmanned aircraft systems-based remote sensing. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8 (5):1876–1885.
- Yu, W. 2004. Practical anti-vignetting methods for digital cameras. *IEEE Transactions on Consumer Electronics* 50 (4):975–983.

Interaction Relationship Between Built-Up Land Expansion and Demographic-Social-Economic Urbanization in Shanghai-Hangzhou Bay Metropolitan Region of Eastern China

Rui Xiao, Xin Huang, Weixuan Yu, Meng Lin, and Zhonghao Zhang

Abstract

Employing coupling coordination degree model (CCDM), this research attempts to reveal the interaction relationships between built-up land expansion (BLE) and demographic-social-economic (DSE) urbanization in Shanghai-Hangzhou Bay (SHB) Metropolitan Region. It first compared the development trends of four elements, including built-up land expansion level, demographic urbanization, social urbanization, and economic urbanization from 1994 to 2015 through descriptive statistics. Then CCDM was used to identify the spiral escalation trend relationships between built-up land expansion and demographic/social/economic urbanization, respectively. The findings revealed that the degree of coupling coordination between BLE and DSE urbanization had the trend to ascend in SHB, and Shanghai has a more superior balanced development tendency than other cities. It concludes that CCDM can be implemented as an effective approach to evaluate the coupling relationship, and the related agencies in SHB can strengthen the coordination to provide suggestions and make decisions for the coordinated development of urban agglomeration.

Introduction

China has been experiencing a rapid urbanization process, which is characterized by strong built-up expansion, extreme urban population growth, and unprecedented social-economic development since the 1990s. During the last several decades, urbanization emerged in the developed areas, where urban construction and development were vigorous, and land use in these regions experienced significant changes and restructuring. Under the force of administrative measure and land finance, China's urbanization is in a stage of fast-forward and uncontrolled spatial sprawl, leading to illusory high population urbanization (Qi *et al.* 2016), high housing prices (Wang *et al.* 2017a; Li *et al.* 2017), disorder expansion of built-up lands (Li *et al.* 2014), environmental degradation (Wang *et al.* 2017b), huge farmland encroachment (Chien 2015), increasing conflicts between human and land (Ma *et*

al. 2016), and so on. Especially the cities in the coastal economic zone, which are the preferred destination for millions of internal migrants and overseas investors, have been the locomotives in propelling China's economic growth (Chen *et al.* 2000; Wu and Zhang 2012). However, the unbalanced disposition of human and land in coastal zones may threaten the healthy development of urbanization. How to coordinate the relationships among built-up land expansion, population growth, and social-economic development is one of the most important issues that China should face during the new urbanization process.

Urbanization involves an array of interactive processes such as demographic, social, and economic. How China's urbanization has been influenced by the built-up land expansion in the eastern coastal region is certainly a significant issue that needs to be explored and addressed. Related existing research is largely dedicated to identifying the urbanization's impacts through the aspects of spatiotemporal changes of built-up land expansion (Chuai *et al.* 2015; Ianos *et al.* 2016) and driving forces and their interactions of built-up land expansion during the urbanization process (Ju *et al.* 2016; Liu *et al.* 2017). Others built models to analyze the interaction relationship between population growth and urban expansion (Marshall 2007; Deng *et al.* 2008). However, comprehensive studies on the coupling coordination relationships among expansion of built-up land, planning policies, and demographic-social-economic urbanization in the metropolitan regions are relatively scarce.

The concept of coupling coordination is oriented from physics, which refers to two or more systems that may influence one another through a variety of interactions (Li *et al.* 2012). Coupling coordination degree was employed to analyze the interaction within systems, since it can reflect to what extent the development of systems is coherent and harmonious when comparing its value of two different subsystems. Coupling coordination model has been applied in geographic studies, such as urbanization and environment (Fang and Wang 2013; Guo *et al.* 2015; Wang *et al.* 2014), economics (Lu *et al.* 2017), tourism and environment (Tang 2015), urbanization and population (Tang *et al.* 2017), and so on. The interaction between the two systems can promote a regional coupling coordination degree, which shows a spiral escalation trend (Lu *et al.* 2017).

Taking the Shanghai-Hangzhou Bay Metropolitan Region (SHB), a region covering the most developed cities in

Rui Xiao, Xin Huang, Weixuan Yu, and Meng Lin are with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China; rxiao@whu.edu.cn (R. X.); xhuang@whu.edu.cn (X. H.); wxyu@whu.edu.cn (W. Y.); mlin_@whu.edu.cn (M. Lin).

Xin Huang is also with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China.

Zhonghao Zhang is with the Institute of Urban Studies, Shanghai Normal University, Shanghai 200234, China.; zzh87@shnu.edu.cn (Z. Z.)

Photogrammetric Engineering & Remote Sensing
Vol. 85, No. 3, March 2019, pp. 231–240.
0099-1112/18/231–240

© 2019 American Society for Photogrammetry
and Remote Sensing
doi: 10.14358/PERS.85.3.231

eastern China as an example, this study explores the status and process of built-up land expansion there, as well as its correlations to demographic, social, and economic urbanization through the coupling coordination degree model. It first compared the development trends of four elements—“built-up land expansion”, “demographic urbanization”, “social urbanization”, and “economic urbanization”—in SHB across the past 21 years (from 1994 to 2015) through descriptive statistics. Then coupling coordination degree model was employed to identify the interactive effects of built-up land expansion and demographic-social-economic (DSE) urbanization. Specifically, this research focuses on (1) spatiotemporal patterns of built-up land expansion (BLE) and DSE urbanization from 1994 to 2015; (2) coupling coordination degree of BLE and DSE urbanization; and (3) implications for coordinated development in SHB.

Study Area

The SHB metropolitan region consists of one core megacity, Shanghai, as well as five cities (Hangzhou, Ningbo, Shaoxing, Jiaxing and Huzhou) in the Zhejiang Province (Figure 1). The SHB covers a total land area of more than 50 000 km² with a total population of more than 38 million (2016 Statistical Yearbook of China). SHB is located on the southern part of Yangtze River Delta, which is the first pole of China’s economic growth. Its climate is humid, subtropical, characterized by short cold dry winters and long, hot, rainy summers. The average annual temperature is 16.4°C, with annual rainfall of 1460 mm. Landforms are diverse, including hills, plateaus, basins and plains.

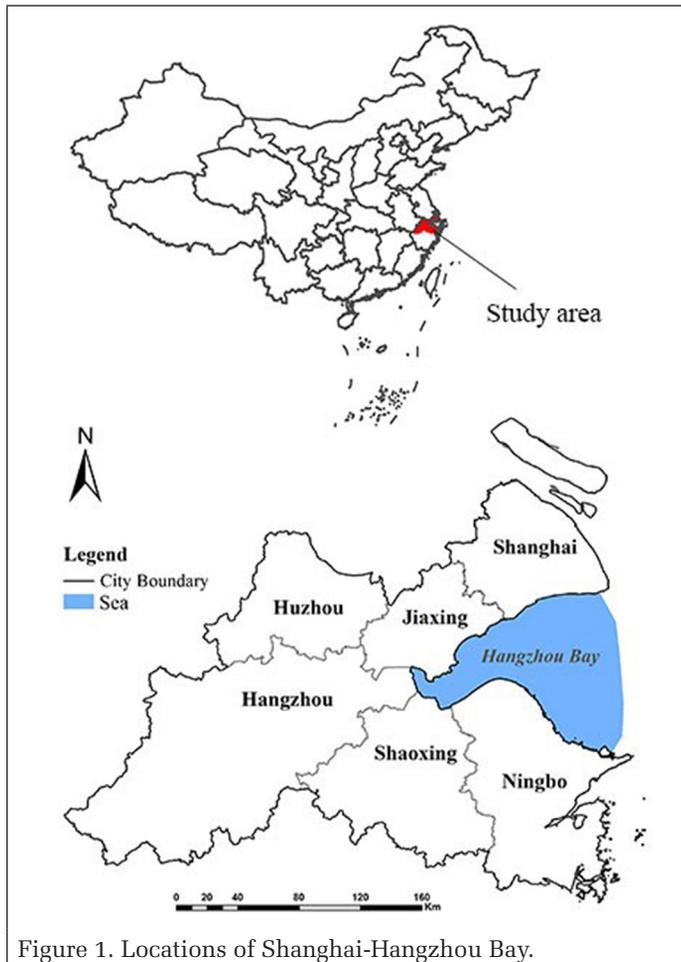


Figure 1. Locations of Shanghai-Hangzhou Bay.

As one of the most important components in the southern part of world-class city groups in the Yangtze River Delta, SHB is a pivotal economic development benchmark in China’s southeast coastal area, as well as a typical manifestation of the rapid development of urbanization in China. China’s market transition since 1994 has pushed SHB into a stage of rapid growth. It has experienced unprecedented economic development and population growth in recent decades. Gross Domestic Product (GDP) per capita was 55 840 renminbi (RMB) in 1994, and reached about 660 890 RMB in 2015, which increased more than 11-fold. In addition, its population density amounted to 683 and 750 persons per km² in 1994 and 2015, respectively, making it one of the most densely populated regions in China. The demand for built-up lands increased rapidly with the population explosion and economic inflation. Therefore, scientific evidence was needed to inform the local government about the relationship between DSE and BLE, and the subsequent rational utilization of built-up lands. SHB was used as a case to evaluate the interactive effects between DSE and BLE in this study.

Materials and Methods

Data Preprocessing

Built-Up Land Detection

Time series of satellite data, including the datasets of multi-spectral Landsat Thematic Mapper (TM)/Enhanced Thematic Mapper Plus (ETM+) (Path 120, Row 38–40; Path 119, Row 38–40; Path 118, Row 38–40) imagery acquired on 1994, 2003, 2009, and 2015, were selected for this study. Prior to interpretation, all images were standardized to the same reference spectral characteristics by atmospheric correction. Then, the images were geometrically rectified to Universal Transverse Mercator (UTM) 50 World Geodetic System 1984 (WGS 84) coordinate system using the quadratic method. All the images were resampled to 30 m using the nearest neighbor algorithm to maintain the unchanged original brightness values of pixels, and the root mean squared error (RMSE) were both calculated within 0.5 pixels. The image processing and data manipulation were conducted using Data Preparation Module supplied with the Environmental Systems Research Institute (ESRI) ERDAS IMAGINE 2014 image processing software. The methods for visual interpretation were adapted from Herold *et al.* (2003) and Munsu *et al.* (2010). The working window was set at a 1:50 000 scale and then built-up boundaries were delineated. Google Earth was used to check the accuracy of image interpretation. For each image, 300 samples were randomly selected to check the accuracy of the built-up land maps. The overall accuracy of built-up land maps between 1994 and 2015 was determined to range from 82.6% to 86.4% with an average of 84.5%, and the Kappa coefficient ranged from 0.81 to 0.84 with an average of 0.82. Then the overlay module in *ARCGIS* 10.1 was used to map built-up land expansion from 1994–2015 (Figure 2).

Statistical Indicators

A database of three important aspects of urbanization—demographic urbanization, social urbanization and economic urbanization was established based on data from statistical yearbooks (Shanghai and Zhejiang Municipal Bureau of Statistics, 2016), China’s urban statistical yearbook (China National Bureau of Statistics, 2016), official websites of local governments, and other statistical outlets (1990–2016), spatially regarding city-level administrative bodies of the SHB region as a basic statistical unit. We standardized the data using Equations 1 and 2 and eliminated the influence of dimension, magnitude, and positive and negative orientation:

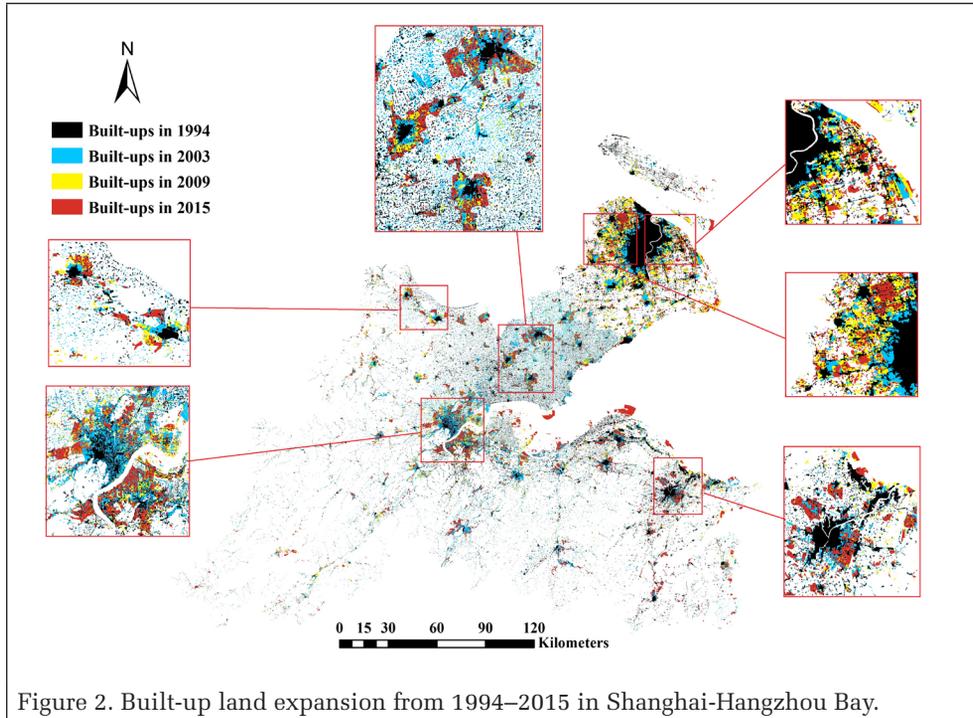


Figure 2. Built-up land expansion from 1994–2015 in Shanghai-Hangzhou Bay.

and availability. Subsequently, three-round Delphi Process and correlation analysis were performed, and a total of 12 indices were generated. Besides, area and percentage of built-up land were used to indicate the built-up land in each year (Table 1).

Evaluation of BLE and DSE Urbanization

The built-up land expansion rate was used to evaluate the spatio-temporal process of built-up land expansion. The expansion rate describes the growth of built-up land area as a percentage of the total built-up land area for a given period. The built-up land expansion annual rate (km^2/year) is given as (Xu and Min 2013):

$$R = \frac{UA_{n+i} - UA_i}{UA_i} \times \frac{1}{n} \times 100\% \quad (3)$$

where R is the urban expansion rate, UA_{n+i} and UA_i are the built-up land area in different time periods $n + i$ and i , respectively, and n is the interval of the whole period (years).

Positive indicator: $r_{ij} = \frac{(x_{ij} - \min(x_j))}{(\max(x_j) - \min(x_j))} \quad (1)$

Negative indicator: $r_{ij} = \frac{(\max(x_j) - x_{ij})}{(\max(x_j) - \min(x_j))} \quad (2)$

where x_{ij} denotes the value of indicator j in year i ; $\max(x_j)$ and $\min(x_j)$ are the maximum value and minimum value of j indicator in all years, respectively. Thus, all the index value will be ranged in $[0,1]$.

Methods

Index System for BLE and DSE Urbanization

Referring to previous conducted studies (Qiao and Fang 2006; Bao and Fang 2007; Li *et al.* 2012), a set of indices focusing on different aspects of demographics, society and economy were initially selected considering of the data integrity, accuracy

Weighting coefficients of all indicators are obtained through mean square error (MSE) decision methods:

Here the normalized data is used to calculate the MSE of random variables of each indicator, which are then normalized for weighting coefficients (Tang *et al.* 2017).

The average value of x' is given as:

$$F(x_i) = \frac{1}{n} \sum_{j=1}^n r_{ij} \quad (4)$$

The MSE of indicator x_i is given as:

$$\zeta(x_i) = \sqrt{\sum_{j=1}^n [r_{ij} - F(x_i)]^2} \quad (5)$$

The weighting coefficient of indicator x_i is given as:

$$\omega_i = \frac{\zeta(x_i)}{\sum_{i=1}^n \zeta(x_i)} \quad (6)$$

Table 1. Index system used for evaluation of BLE and DSE urbanization.

Subsystem	First Grade Index	Basic Grade Indices
The integration value of BLE	Built-up land expansion level	Area of built-up land (km^2)
		Percentage of built-up land (%)
The integration value of DSE urbanization	Demographic urbanization level	Number of total population
		Percentage of non-agricultural population (%)
		Number of social industry employment
		Number of primary industry employment
	Social urbanization level	Number of academic school students per 10 000 people
		Number of hospital/clinic beds per 10 000 people
		Number of libraries per 10 000 people
	Economic urbanization level	Output value of the primary industries (Yuan)
		Total investment in fixed assets (Yuan)
		Per capita income of rural residents (Yuan)
Total retail sales of consumer goods (Yuan)		
		Total power of agricultural machinery (kilowatt)

The evaluation of a single indicator is given as:

$$S_{ij} = \omega_i \times r_{ij} \quad (7)$$

The comprehensive level in year i is given as:

$$S_i = \sum_{j=1}^n S_{ij} \quad (8)$$

where r_{ij} is the value of each indicator after standardization, $F(x_i)$ is the average value of indicator x_i , $\zeta(x_i)$ is the MSE of indicator x_i , ω_i is weighting coefficient of indicator x_i , S_{ij} is the evaluation of a single indicator, and S_i is the comprehensive level in year i .

Coupling Coordination Degree Model (CCDM)

Coupling coordination degree model (CCDM) was employed in this study to reveal the interaction relationship between demographic-social-economic urbanization and built-up land expansion. The variable values were defined as U_0 , U_1 , U_2 , and U_3 , which refer to built-up land expansion level, demographic urbanization, social urbanization, and economic urbanization, respectively.

The CCDM is given as (Illingworth 1996):

$$C = n \left\{ (U_0 \cdot U_1 \cdot U_2 \cdots U_n) / \left[\prod (U_i + U_j) \right] \right\}^{1/n} \quad (9)$$

where U_i represents the contribution of sub-system i to the total system. C is the coupling degree and the value of C ranges from 0 to 1, which reflects the strength of mutual influences among systems from U_1 to U_n . The coupling coordination degree model can further reflect to what extent the development of systems is coherent and harmonious. In this study, we calculated the coupling coordination degree of two sub-systems U_a and U_b , and the formula is given as:

$$D = (C \times T)^{1/2} \text{ and } T = \alpha U_a + \beta U_b \quad (10)$$

where C represents the coupling coordination degree between U_a and U_b , and T reflects the overall effect and level of U_a and U_b , whereas α and β denote undetermined coefficients, which are 0.5 in this study because U_a and U_b are equally important. The coupling coordination degree D ranges from 0 to 1, whereas higher value indicates higher coherence level among sub-systems. The coupling coordination periods was divided into four distinct stages according to the previous references on coupling coordination division standard (Tang 2015; Sun *et al.* 2016; Ai *et al.* 2016), including low level coupling, antagonistic period, running-in period, and high-level coupling. When $0 \leq D < 0.2$, the coordination of BLE and DSE is in the low-level coupling stage, in which there is little influence between each

other. When $0.2 \leq D < 0.5$, the coupling coordination of BLE and DSE is in the antagonistic period, in which built-up land expansion would restrict the DSE urbanization. When $0.5 \leq D < 0.8$, the coordination of BLE and DSE is in the running-in period, in which the relationship between BLE and DSE gradually become more optimized. When $0.8 \leq D \leq 1.0$, the coupling coordination of BLE and DSE is in the high-level coupling stage, in which the relationship between BLE and DSE is coordinated optimization, and each aspect would propel the development of the other aspect. Besides, there are two kinds of relationship between BLE and DSE in different stages: BLE development is ahead of or lags behind DSE urbanization. Therefore, the four stages were divided into eight intervals according to the relationships between BLE and DSE in different stages (Figure 3). The explanation of all the intervals are shown in Table 2.

Results

Evaluation of BLE and DSE Urbanization

Table 3 shows the expansion patterns of the built-up land of the SHB during different periods, which can be described as three stages. (1) During 1994 and 2003, Shanghai had the lowest built-up land expansion rate, whereas the cities in Zhejiang Province had a rapid expansion in this time, especially for Hangzhou, Jiaying and Shaoxing, which witnessed an expansion rate increase more than 15%. In this stage, the built-up land areas in Shanghai reached from 1350 km² to

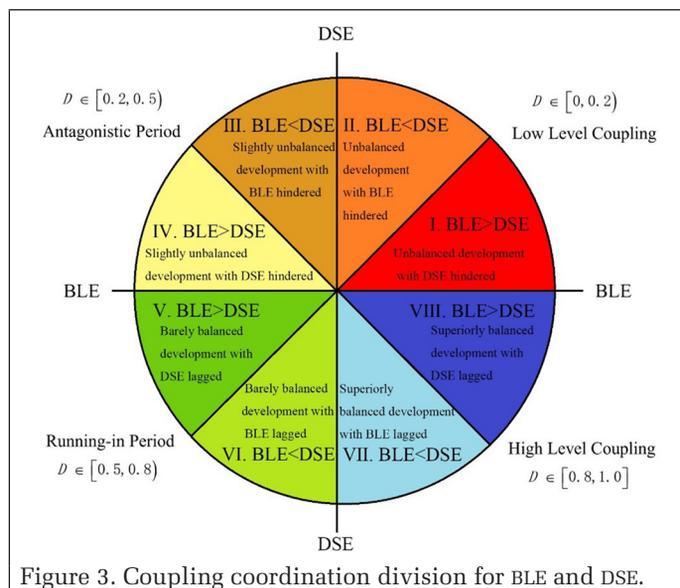


Figure 3. Coupling coordination division for BLE and DSE.

Table 2. Explanation for coupling coordination division.

D Value	Division	U_{BLE} and U_{DSE}	Stage	Interval	Features
[0,0.2)	Unbalanced Development	$U_{BLE} > U_{DSE}$	Low level coupling	I	Unbalanced development with DSE hindered, in which the DSE has little influence on the BLE development.
		$U_{BLE} < U_{DSE}$	Low level coupling	II	Unbalanced development with BLE hindered, in which the BLE has little influence on the DSE development.
[0.2,0.5)	Slightly Unbalanced Development	$U_{BLE} < U_{DSE}$	Antagonistic Period	III	Slightly unbalanced development with BLE hindered, in which the BLE would restrict the DSE development.
		$U_{BLE} > U_{DSE}$	Antagonistic Period	IV	Slightly unbalanced development with DSE hindered, in which the DSE would restrict the BLE development.
[0.5,0.8)	Barely Balanced Development	$U_{BLE} > U_{DSE}$	Running-in Period	V	Barely balanced development with DSE lagged, in which DSE has ample room for development in a relatively stable subsystem.
		$U_{BLE} < U_{DSE}$	Running-In Period	VI	Barely balanced development with BLE lagged, in which BLE would develop more rapidly to catch up with the DSE development.
[0.8,1.0)	Superiorly Balanced Development	$U_{BLE} < U_{DSE}$	High Level Coupling	VII	Superiorly balanced development with BLE lagged, in which DSE urbanization is in a mature stage that can propel built-up land expansion in a stable subsystem.
		$U_{BLE} > U_{DSE}$	High Level Coupling	VIII	Superiorly balanced development with DSE lagged, in which built-up land expand ahead that can provide great room for DSE development in a stable subsystem.

Table 3. Built-up land expansion rate in the SHB (Unit: %).

Period	1994–2003	2003–2009	2009–2015	1994–2015
Shanghai	3.5	6.3	3.1	5.5
Hangzhou	15.9	3.4	5.4	13.7
Ningbo	5.4	1.9	5.8	5.9
Jiaxing	16.2	1.5	2.8	10.1
Huzhou	13.1	3.4	5.1	11.5
Shaoxing	16.6	1.6	4.9	12.1
Shanghai-Hangzhou Bay	8.6	3.5	4.1	8.0

1770 km², and the percentage increased from 20% to 26%, lagged behind Jiaxing, which saw the percentage increased from 11% to 28% (Figure 4). (2) During 2003 and 2009, Shanghai revealed highest values in built-up expansion rate, which displayed a new sprawl boom in this period. However, the cities in Zhejiang Province slowed the pace of expansion, with the percentage fluctuating between 1% and 3% (Figure 4). (3) During 2009 and 2015, Hangzhou, Ningbo, Huzhou, and Shaoxing had an expansion rate around 5%, almost as the same value as in Shanghai, showing that these cities began a new growth under the influence of Shanghai. For the whole Shanghai-Hangzhou Bay, built-up land increased by 5320 km², from 3208.4 km² to 8528.4 km², at an annual increase rate of 253.3 km² during the period from 1994 to 2015, which is lower than the period during 1994 and 2003 and higher than the other two periods, indicating that the expansion of built-up land in the SHB mainly occurred in the first stage. Built-up lands increased slower during 2003–2009 than the other two stages in SHB, resulting from that the built-up lands in cities of Zhejiang Province expanded slowly in this period. Synthetically, SHB reached at 8% annual increase from 1994 to 2015, whereas Shanghai and Ningbo lagged behind the whole study area, increasing by 5.5% and 5.9%, respectively.

There was a significant spatial variability between BLE and DSE urbanization. Shanghai and Ningbo saw a lower annual rate of built-up land expansion, suggesting that these two cities had controlled the expansion of built-up lands during 1994 and 2015. Hangzhou and Shaoxing had a higher increase of built-up land expansion with annual rate of more than 12.1%, which indicated that these two cities had a great demand for built-up lands and this need converted large amounts of land into built-up lands, promoting the rapid growth of urban built-up areas.

Demographic urbanization resulted that Jiaxing and Shaoxing had an annual increase rate of more than 1.5%, while Shanghai and Hangzhou followed this number (0.81%–1.50%). Huzhou and Ningbo had demographic urbanization annual rate lower than 0.80%. The results of social

urbanization showed that Shaoxing witnessed the highest average annual rate more than 50%, while Shanghai and Hangzhou had the lowest average annual rate less than 30%. In addition, Hangzhou and Ningbo were in the first rank of economic urbanization annual rate more than 50%, while Shanghai had the lowest number less than 30%.

Overall, Shanghai saw a lower annual rate in all the aspects of built-up land expansion and demographic-social-economic urbanization than most of the cities in Zhejiang Province, which indicated that on the one hand Shanghai had a bigger radix number than other cities and on the other hand the Shanghai government had a good control in urbanization. Shaoxing had a higher annual rate in built-up land expansion, demographic and social urbanization, displaying a rapid development during the two decades. Hangzhou and Ningbo were the two strong economic cities in Zhejiang Province and also showed the dominant economic positions during 1994 and 2015 (Figure 5). Jiaxing and Shaoxing had the highest annual rate of demographic urbanization, representing that people would consider the factors of urban construction and social urbanization when they decided to migrate to other cities. Especially for Shaoxing, which had the most rapid development rates of built-up lands and social urbanization, and accordingly absorb most people to move into the city for urban construction.

Comprehensive Levels in the BLE and DSE Urbanization

Demographic urbanization in Shanghai witnessed the most rapid development, while the built-up land expansion rate cannot catch demographic and social urbanization. It indicates that more built-up lands should be developed to satisfy the demands of housing and social activities for human beings. There was a steady increasing trend for all the aspects during 1994 and 2003 in Hangzhou, and an enormous increasing trend for economic and social urbanization from 2009 to 2015, both reaching more than 0.8. Ningbo saw the most rapid development in economic urbanization, far ahead of other aspects, which demonstrated that built-up land expansion should be emphasized on economic construction. Jiaxing had a larger range of built-up land expansion, while other aspects were in a steady developing trend. It suggested that there existed enough built-up land areas for DSE urbanization in Jiaxing. Levels of built-up land expansion in Huzhou lagged behind demographic-social-economic urbanization in 1994 and 2015, suggesting that built-up lands cannot meet the demand of DSE urbanization. Shaoxing had the most development of economic urbanization and the least development of built-up land expansion, indicating that Shaoxing focused more on the economic development and ignored the construction for built-up lands.

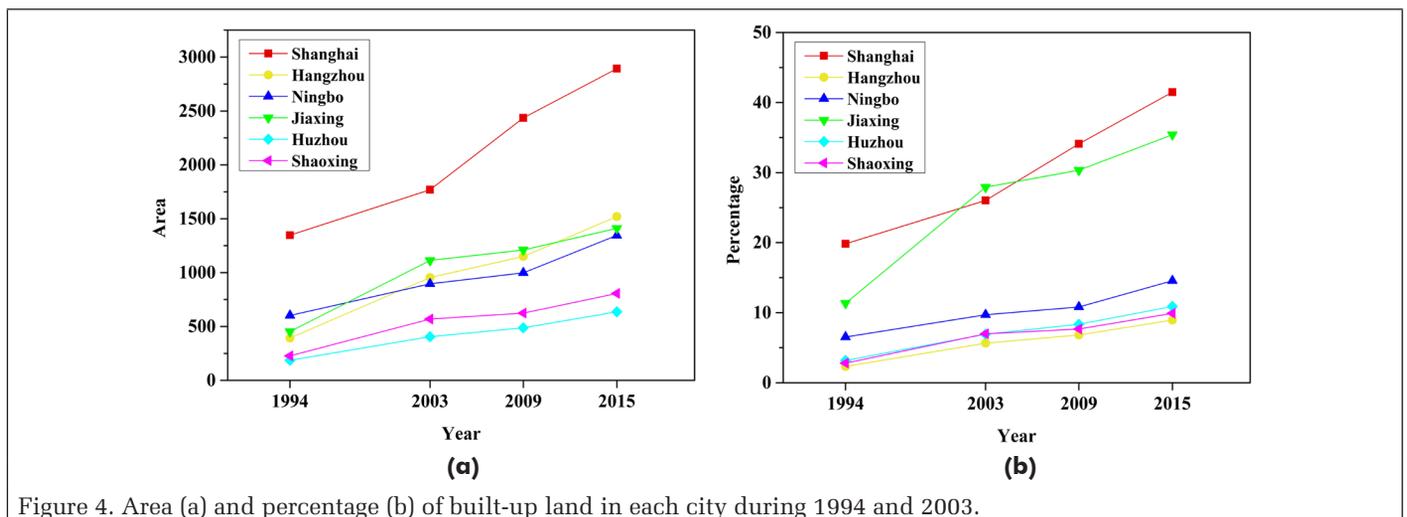


Figure 4. Area (a) and percentage (b) of built-up land in each city during 1994 and 2003.

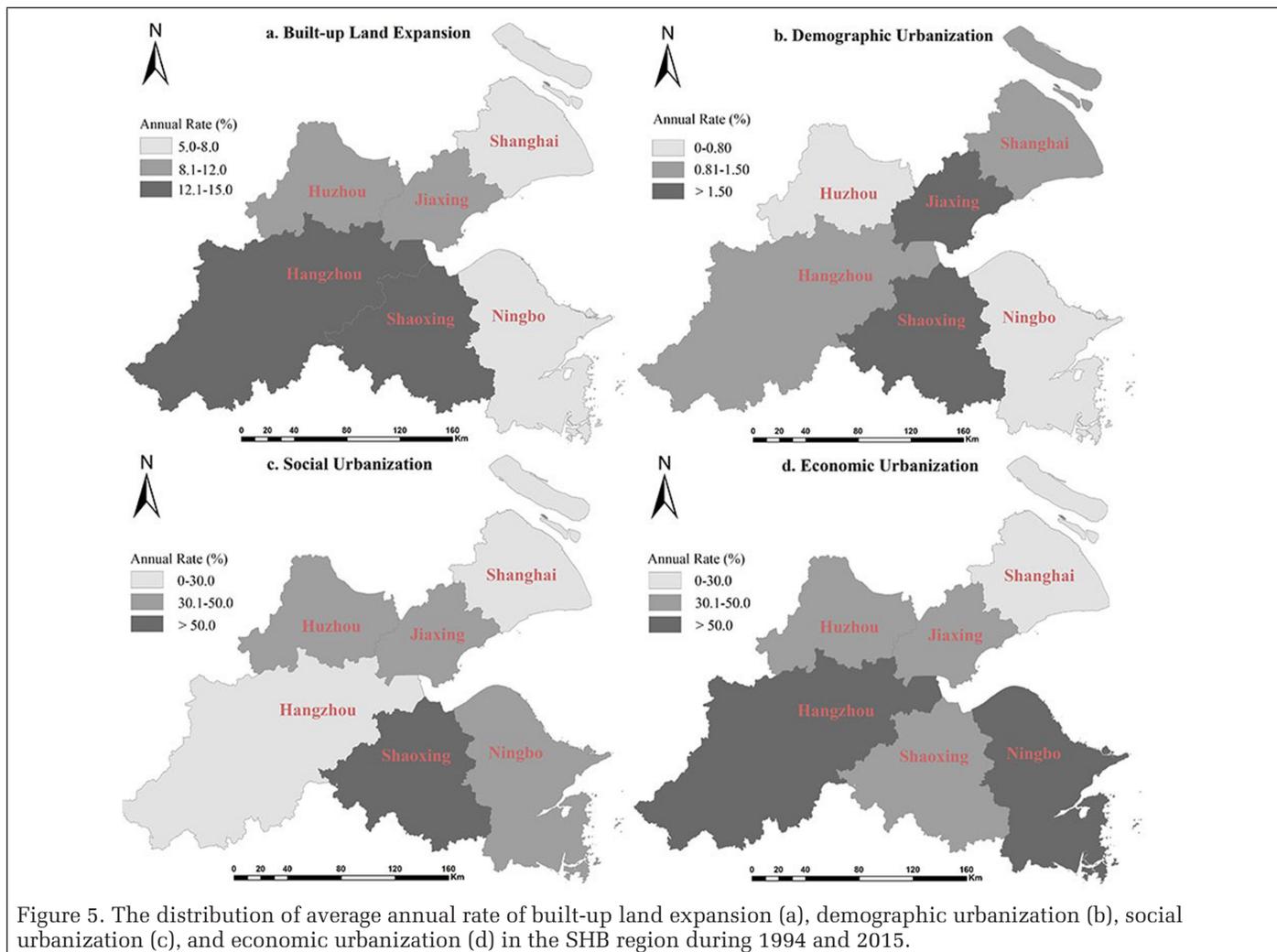


Figure 5. The distribution of average annual rate of built-up land expansion (a), demographic urbanization (b), social urbanization (c), and economic urbanization (d) in the SHB region during 1994 and 2015.

Coupling Coordination Degree of BLE and DSE Urbanization

Figure 6a shows that many cities are in III and IV intervals for the coupling coordination of BLE and demographic urbanization in 1994, which means a slightly unbalanced stage. Most cities reached barely balanced stage in 2015, which are V and VI Intervals, respectively. The development of Shanghai was ahead of other cities, reaching VI interval in 1994 and 2003 and VII interval in 2009 which is a stage of superiorly balanced development with BLE lagged. It indicated that the built-up expansion in 2015 provided sufficient space for demographic urbanization development; in other words, Shanghai has plenty of built-up land areas for more immigrants. This result was related to the population policy management issued by Shanghai, where there were many restrictions for people to obtain the household registration. Therefore, as the first-tier city and one of the fastest economically growing cities in China, Shanghai witnessed a nonrapid growth of the population. This well explained the reason that built-up lands provided enough spaces for population growth. Jiaxing lagged behind other cities in the coupling coordination of BLE and demographic urbanization, which was in barely balanced development with demographic urbanization lagged in 2015. Jiaxing had more intensity of built-up land expansion than demographic-social-economic urbanization (Figure 7), but the coupling-coordination result reported that Jiaxing had not attracted enough people to balance the coupling development of built-up land expansion and demographic urbanization. Hangzhou and Ningbo had a similar development trend, which was in IV interval at the beginning, and then came into

the V interval and VI at last. It displayed that these two cities had a stable development for coupling coordination of BLE and demographic urbanization.

Figure 6b shows the coupling coordination of BLE and social urbanization. Shanghai was still ahead of other cities. The development was from V interval in 1994 to VI interval in 2003 and finally to VII interval in 2009 and 2015, which was in the stage of superiorly balanced development with BLE lagged, indicating that the built-up land had already provided enough space for social urbanization in 2009. Hangzhou, Ningbo and Jiaxing were in IV interval in 1994, which is slightly unbalanced. Among these three cities, Hangzhou had a stable increasing trend in the latter stages of development, reaching VI interval in 2009 and 2015. While Ningbo and Jiaxing were in V interval in 2015, which is barely balanced development with social urbanization lagged, displaying a weak social development in these two cities. Huzhou and Shaoxing had similar developing trend, which were in II and I intervals in 1994, respectively, representing an unbalanced coordination. However, they developed from IV interval in 2009 to VI interval in 2015, skipping the V interval. It indicated that these two cities had emphasized on coordination development between social urbanization and built-up land expansion.

Figure 6c shows that Shanghai changed from IV interval in 1994 to V interval in 2003 and changed from V interval in 2009 to VIII interval in 2015 (skipping the VI and VII intervals), which was in the stage of superiorly balanced development with economic urbanization lagged. It displayed that Shanghai had developed into a very high stage that built-up

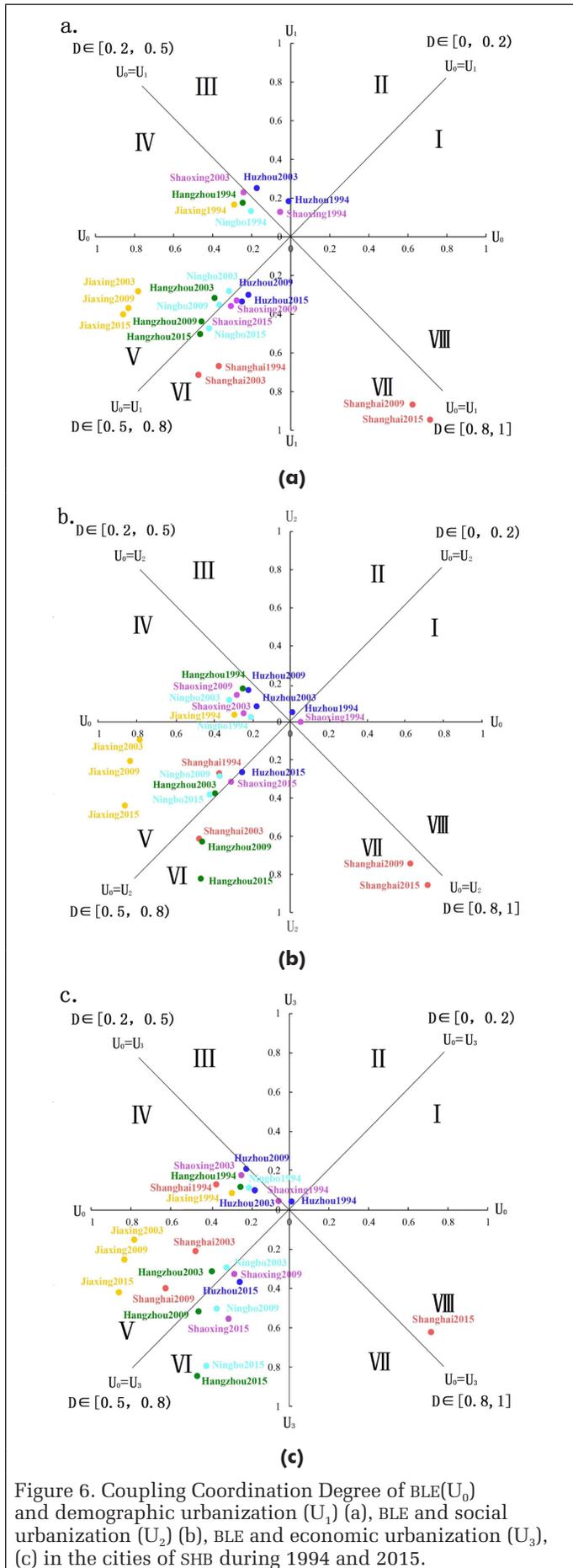


Figure 6. Coupling Coordination Degree of BLE (U_0) and demographic urbanization (U_1) (a), BLE and social urbanization (U_2) (b), BLE and economic urbanization (U_3), (c) in the cities of SHB during 1994 and 2015.

land areas were able to provide enough room for the economic urbanization, showing an emphasis from the government on the coordination development between BLE and economic urbanization. Jiaxing had a similar BLE-Economic Urbanization developing trend as BLE-Social Urbanization, which reached V interval in 2015, indicating a weak coordination of built-up land and social urbanization as well. Hangzhou and Ningbo had similar changing trend during 1994 and 2015, reaching the VI interval at last, which is barely balanced development with BLE lagged. Huzhou had a rapid growth of development during 1994 and 2015. It grew from the unbalanced development stage in 1994 to IV interval in 2003, and from IV interval in 2009 to VI interval in 2015, which showed that the Huzhou government had emphasized on the coordination development between built-up land expansion and economic urbanization.

Discussion

The results of the coupling coordination degree are essential for cities in SHB in their future development planning. It is important to monitor the coupling coordination process and use the coupling coordination model to understand sustainable development between BLE and DSE urbanization in the SHB region. Demographic increase could surely improve social and economic benefits. In turn, the increase benefits of society and economy could absorb more people to move in (Hasan *et al.* 2017). During the cycling process, the expansion trend of built-up lands has been pushed forward. Besides, a great amount of young immigrant population from rural to urban would ask for more houses to live. And with social development, more and more public basic facilities, such as healthy communities, cultural facilities, schools, recreational facilities, art museums, gymnasiums, etc., were needed to satisfy people's demands of material culture and spiritual culture. Besides, rapid development of economy also brings increases in investment in urban construction, which leads to urban area to expand at an accelerated rate (Xu *et al.* 2000). All these aspects account for the constant expansion of built-up lands. In addition, built-up land expansion provides strong basic support and construction space for the development of demographic-social-economic urbanization. In another word, when the built-up lands cannot satisfy the demands of socio-economic development, the former would have an inhibitory effect on the latter to some extent. With the population increase, people have great demands for land resources and the land prices increase accordingly, which make the fringe areas become one of the best choices for investors. Developers choose suburbs as residential areas, and factories and schools move to the surrounding areas of the city, all of which lead to the construction of a serious of supporting facilities in these areas such as housing, schools, hospitals, transportation and service facilities (Tian *et al.* 2017). Moreover, the investment of financial business and improvement of social public service facilities can absorb more outlanders pouring into the urbans, and accordingly these people would contribute to the development and construction of those cities, and to cause the built-up land expansion year after year.

The increasing rate of built-up lands in Shanghai lags behind that of other cities in Zhejiang Province; it results from this city having had rapid development in the early 1990s, when some New Districts such as Pudong and Jiading were established in 1992 (Shi *et al.* 2009), so Shanghai displayed a low increasing rate value during after 1994. Then, due to China's Development Zone Policy, which was successful in attracting foreign direct investments and promote regional economic development (Zhang 2011), land price in central urban grows rapidly. This impelled many medium-sized and small enterprises to open factories in villages or towns, which may absorb the rural people who were restricted by the household

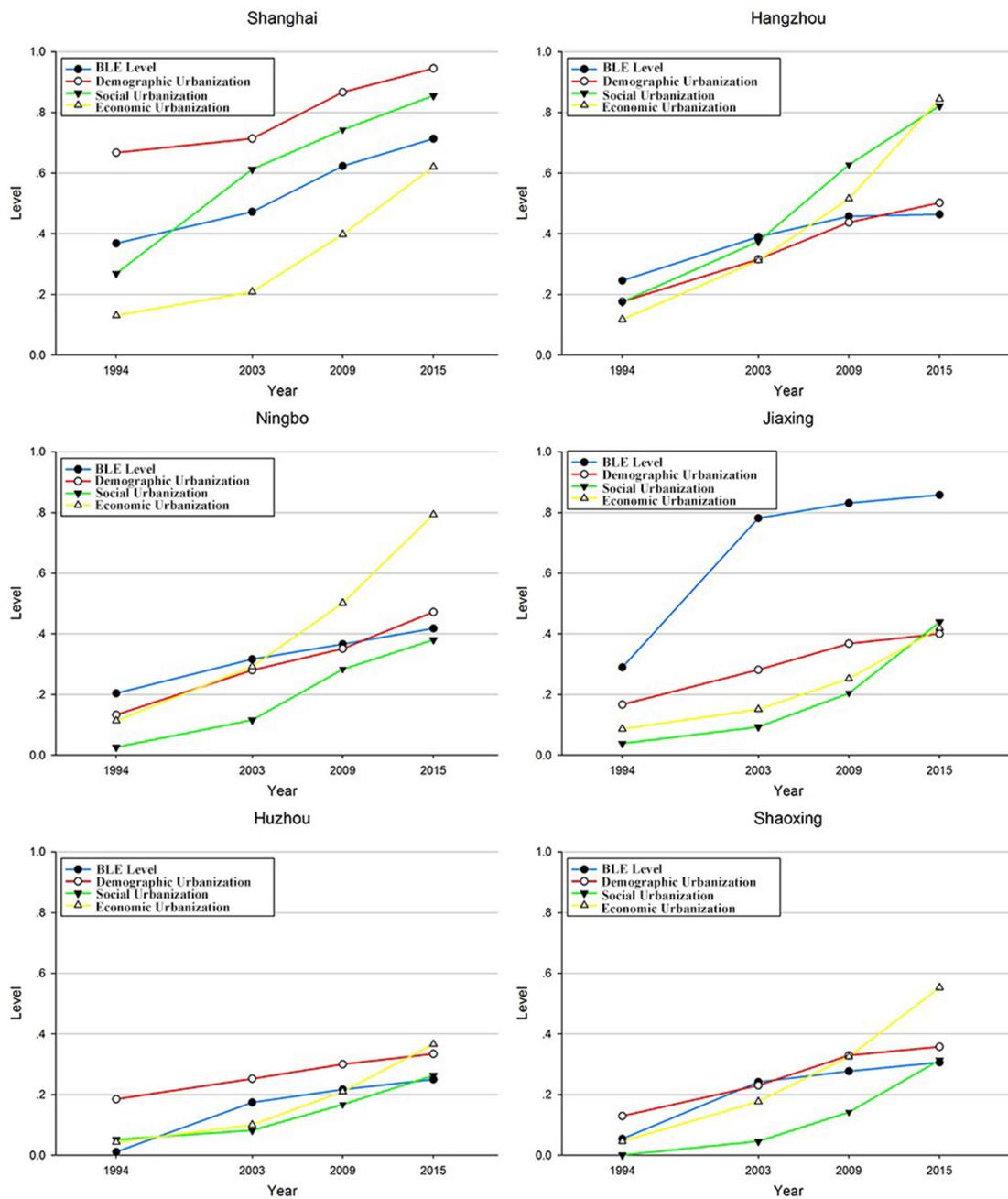


Figure 7. Trends of comprehensive levels in the BLE and DSE urbanization subsystem.

registration or cannot afford the cost living in central urban. Therefore, it significantly resulted in the accelerated population migration from rural to other remote areas to the developed villages or towns near the cities and sped up the built-up land expansion as well (Wu and Zhang 2012). Under these policies, built-up land expansion showed a rapid development from 1994–2003. Then in 2004, the central government issued a policy on strict land management system, aiming to control construction land growth and prevent cultivated land loss, in order to restrain the excessive land expropriation in the urban fringe to avoid or mitigate farmland degradation (Liu *et al.* 2015). Therefore, in our research, SHB displayed the lowest annual increase rate during 2003–2009 than other two stages.

Propelled by industrialization and urbanization, there is a great need of land resources for socio-economic development, pushing forward the built-up land expansion and absorbing a great amount of population to accumulate. Nevertheless,

because of the certain lag comparing urban construction to population attraction and the restriction by the household registration system, the expansion rate of built-up lands would show faster trend than population growth. With the implementation of China's Yangtze River Economic Belt strategy, the social and economic development of urban agglomeration centered on Shanghai and Hangzhou will continue to be in the steady development trend. In addition, since urbanization was a key national development strategy (Bloom *et al.* 2008; Siciliano 2012), the urbans still demand for enormous land resources, which brings great challenges to the coordinated development of city, people and land.

Currently, the high speed of expansion of built-up lands and development of demographic-social-economic urbanization attracts a great deal of attention due to the extreme population boom, severe environmental damage and scarcity of land resources. By collaboration with adjacent cities,

governments introduced a series of policies to monitor land use and urban planning, as well as a great number of decisions, such as household registration policy (Tyner and Ren 2016), housing restriction measures (Zhou 2016), agricultural tax reform (Takeuchi 2013), etc. to limit the freedom of rural population migration into urbans, aiming at avoiding uncontrolled demographic urbanization and built-up land expansion which may cause adverse social, land and environmental issues (Henríquez *et al.* 2006; Dewan and Yamaguchi 2009; Abu Hammad and Tumeizi 2012; Siciliano 2014). Since built-up land expansion will inevitably occupy large number of natural and seminatural lands, which can be the most intensive disturbance to natural ecosystems (Qin *et al.* 2017). Shen *et al.* (2017) found that megacities such as Beijing and Shanghai experienced increases in particulate matter (PM) 2.5 exposure due to migrants swarming into cities and rapid urbanization. Cao *et al.* (2017) pointed out that urbanization would cause haze pollution, which could further enhance urban heat. Xie *et al.* (2017) reported that land use change, especially urban expansion, has caused lake degradation and water pollution in the middle and lower Yangtze floodplain. Besides, built-up land expansion can also lead to soil sealing (Xiao *et al.* 2013), soil degradation (Tefaye *et al.* 2015; Khaledian *et al.* 2017), green space reduction (Chan and Vu 2017; Liang *et al.* 2017), agricultural abandonment and intensification (Detsis 2010) and so on. These impacts on the eco-environment should all be seriously considered in the future in SHB region. Obviously, the interaction of different cities in urban metropolitan areas will become more and more frequent, and the population, society and economy will be closely connected. The environmental and ecological impacts by the built-up land expansion in one city are likely to cause harm to the environment and ecology in neighboring cities, which would indirectly influence social and economic development. Therefore, coordination among the neighboring governments in SHB should be strengthened to balance the conflicts between built-up land expansion and social-economic development.

The objective results could help identify contributions of BLE and DSE urbanization in subsystems and understand the complicated coupling coordination relationship, and then implement the coupling development policies to better balance built-up land expansion and demographic-social-economic urbanization. Obviously, with China's economy entering the new normal and China's strategic planning of the Yangtze River Economic Belt, the social and economic development of the Yangtze River Delta is swift and intense. The urban construction is still urgent for land resources in the future, which will promote further expansion of built-up lands, causing serious challenges to coordinated development of people and land. Consequently, further studies are essential to establish dynamic simulation for coordinated development of built-up land expansion and population growth, when considering how to change the mode of land resources exploitation and utilization, to control the excessive urban land expansion effectively and absorb the demographic accumulation reasonably.

Conclusions

Using SHB as an example, this study employed CCDM to analyze the changing trend of six cities, and to identify their developing process during 1994–2015 for providing suggestions for the future development. The harmonious development of BLE and DSE urbanization system is a dynamic process, which can be quantitatively evaluated by coupling coordination degree model. At different stages, built-up land expansion could produce push or pull forces toward demographic-social-economic urbanization. In turn, urban development also has feedback influences on built-up land expansion

at the same time. This is a nonlinear interaction that can be identified through CCDM. As the above result shows, Shanghai witnessed a lower annual rate of built-up land expansion and social-economic urbanization, while this city has a superior coupling coordination degree due to its successfully implemented policies and land planning. In addition, the degree of coupling coordination between BLE and DSE had the trend to ascend in general during 1994 and 2015, indicating that most cities in SHB emphasized the coordinate development between BLE and DSE, while there are still some cities such as Jiaying that ignored this issue. Consequently, the government and decision makers should pay particular attention to monitoring the coupling subsystems and understanding the factors influencing the degrees of coupling coordination between built-up land expansion and demographic-social-economic urbanization. Besides, taking advantage of superiorly balanced development of Shanghai, all the governments should strengthen the coordination to provide suggestions for the development of urban agglomeration in SHB.

Acknowledgments

The research was supported by the National Natural Science Foundation of China under Grants 41701484, and 41771360, and the Hubei Provincial Natural Science Foundation of China under Grant 2017CFA029.

References

- Hammad, A. A. and A. Tumeizi. 2012. Land degradation: Socioeconomic and environmental causes and consequences in the eastern Mediterranean. *Land Degradation & Development* 23: 216–226.
- Ai, J., L. Feng, X. Dong, X. Zhu and Y. Li. 2016. Exploring coupling coordination between urbanization and ecosystem quality (1985–2010): A case study from Lianyungang City, China. *Frontiers of Earth Science* 10: 527–545.
- Bao, C. and C. L. Fang. 2007. Water resources constraint force on urbanization in water deficient regions: A case study of the Hexi Corridor, arid area of NW China. *Ecological Economics* 62: 508–517.
- Bloom, D. E., D. Canning and G. Fink. 2008. Urbanization and the wealth of nations. *Science* 319: 772–775.
- Cao, C., X. Lee, S. Liu, N. Schultz, W. Xiao, M. Zhang and L. Zhao. 2016. Urban heat islands in China enhanced by haze pollution. *Nature Communications* 7: 12509.
- Chan, K. M. and T. T. Vu. 2017. A landscape ecological perspective of the impacts of urbanization on urban green spaces in the Klang Valley. *Applied Geography* 85: 89–100.
- Chen, S., S. Zeng and C. Xie. 2000. Remote sensing and GIS for urban growth analysis in China. *Photogrammetric Engineering & Remote Sensing* 66: 593–598.
- Chien, S. S. 2015. Local farmland loss and preservation in China: A perspective of quota territorialization. *Land Use Policy* 49: 65–74.
- Chuai, X., X. Huang and Q. Lu. 2015. Spatiotemporal changes of built-up land expansion and carbon emissions caused by the Chinese construction industry. *Environmental Science and Technology* 49: 13021–13030.
- Deng, X. Z., J. K. Huang, S. Rozelle and E. Uchida. 2008. Growth, population and industrialization, and urban land expansion of China. *Journal of Urban Economics* 63: 96–115.
- Detsis, V. 2010. Placing land degradation and biological diversity decline in a unified framework: Methodological and conceptual issues in the case of the north Mediterranean region. *Land Degradation & Development* 21: 413–422.
- Dewan, A. M. and Y. Yamaguchi. 2009. Land use and land cover change in Greater Dhaka, Bangladesh: Using remote sensing to promote sustainable urbanization. *Applied Geography* 29: 390–401.
- Fang, C. and J. Wang. 2013. A theoretical analysis of interactive coercing effects between urbanization and eco-environment. *Chinese Geographical Science* 23: 147–162.

- Guo, Y., H. Wang, P. Nijkamp and J. Xu. 2015. Space-time indicators in interdependent urban-environmental systems: A study on the Huai River Basin in China. *Habitat International* 45:135–146.
- Hasan, S., X. Wang, Y. Khoo and G. Foliente. 2017. Accessibility and socio-economic development of human settlements. *PLoS ONE* 12: e0179620.
- Henríquez, C., G. Azócar and H. Romero. 2006. Monitoring and modeling the urban growth of two mid-sized Chilean cities. *Habitat International* 30: 945–964.
- Herold, M., N. C. Goldstein and K. C. Clarke. 2003. The spatiotemporal form of urban growth: Measurement, analysis and modeling. *Remote Sensing of Environment* 86: 286–302.
- Ianos, I., I. Sirodoev, G. Pascariu and G. Henebry. 2016. Divergent patterns of built-up urban space growth following post-socialist changes. *Urban Studies* 53: 3172–3188.
- Illingworth, V. 1996. *The Penguin Dictionary of Physics*. 92–93. Beijing, China: Foreign Language Press.
- Ju, H., Z. Zhang and L. Zuo. 2016. Driving forces and their interactions of built-up land expansion based on the geographical detector—A case study of Beijing, China. *International Journal of Geographical Information Science* 30: 2188–2207.
- Khaledian, Y., F. Kiani, S. Ebrahimi, E. C. Brevik and J. Aitkenhead-Peterson. 2017. Assessment and monitoring of soil degradation during land use change using multivariate analysis. *Land Degradation & Development* 28: 128–141.
- Li, S., X. Ye and J. Lee. 2017. Spatiotemporal analysis of housing prices in China: A big data perspective. *Applied Spatial Analysis and Policy* 10: 421–433.
- Li, Y., Y. Li, Y. Zhou, Y. Shi and X. Zhu. 2012. Investigation of a coupling model of coordination between urbanization and the environment. *Journal of Environmental Management* 98: 127–133.
- Li, Y., Y. Shi, X. Zhu, H. Cao and T. Yu. 2014. Coastal wetland loss and environmental change due to rapid urban expansion in Lianyungang, Jiangsu, China. *Regional Environmental Change* 14: 1175–1188.
- Liang, H. L., W. Z. Li, Q. P. Zhang, W. Zhu, D. Chen, J. Liu and T. Shu. 2017. Using unmanned aerial vehicle data to assess the three-dimension green quantity of urban green space: A case study in Shanghai, China. *Landscape and Urban Planning* 164: 81–90.
- Liu, T., H. Liu and Y. Qi. 2015. Construction land expansion and cultivated land protection in urbanizing China: Insights from national land surveys, 1996–2006. *Habitat International* 46: 13–22.
- Liu, W., J. Liu and W. Kuang. 2017. Examining the influence of the implementation of major function-oriented zones on built-up area expansion in China. *Journal of Geographical Sciences* 27: 643–660.
- Lu, H., L. Zhou, Y. Chen, Y. An and C. Hou. 2017. Degree of coupling and coordination of eco-economic system and the influencing factors: A case study in Yanchi County, Ningxia Hui Autonomous Region, China. *Journal of Arid Land* 9: 446–457.
- Ma, R., T. Wang, W. Zhang, J. Yu, D. Wang, L. Chen, Y. Jiang and G. Feng. 2016. Overview and progress of Chinese geographical human settlement research. *Journal of Geographical Sciences* 26: 1159–1175.
- Marshall, J. D. 2007. Urban land area and population growth: A new scaling relationship for metropolitan expansion. *Urban Studies* 44: 1889–1904.
- Munsi, M., S. Malaviya, G. Oinam and P. K. Joshi. 2010. A landscape approach for quantifying land-use and land-cover change (1976–2006) in middle Himalaya. *Regional Environmental Change* 10: 145–155.
- Qi, W., S. Liu, M. Zhao and Z. Liu. 2016. China's different spatial patterns of population growth based on the “Hu Line”. *Journal of Geographical Sciences* 26: 1611–1625.
- Qiao, B. and C.L. Fang. 2006. Study on the interactive intimidating relation between urbanization and environment in arid area based on grey system theory. *IEEE International Geoscience and Remote Sensing Symposium* 16: 3409–3414.
- Qin, Y., X. Xiao, J. Dong, B. Chen, F. Liu, G. Zhang, Y. Zhang, J. Wang and X. Wu. 2017. Quantifying annual changes in built-up area in complex urban-rural landscapes from analyses of PALSAR and Landsat images. *ISPRS Journal of Photogrammetry and Remote Sensing* 124: 89–105.
- Shen, H., S. Tao, Y. Chen, P. Ciais, B. Güneralp, M. Ru, Q. Zhong, X. Yun, X. Zhu, T. Huang, W. Tao, Y. Chen, B. Li, X. Wang, W. Liu, J. Liu and S. Zhao. 2017. Urbanization-induced population migration has reduced ambient PM_{2.5} concentrations in China. *Science Advances* 3: e1700300.
- Shi, L., G. Shao, S. Cui, X. Li, T. Lin, K. Yin and J. Zhao. 2009. Urban three-dimensional expansion and its driving forces—A case study of Shanghai, China. *Chinese Geographical Science* 19: 291–298.
- Siciliano, G. 2012. Urbanization strategies, rural development and land use changes in China: A multiple-level integrated assessment. *Land Use Policy* 29: 165–178.
- Siciliano, G. 2014. Rural-urban migration and domestic land grabbing in China. *Population Space and Place* 20: 333–351.
- Sun, Y., Y. Cui and H. Huang. 2016. An empirical analysis of the coupling coordination among decomposed effects of urban infrastructure environment benefit: Case study of four Chinese autonomous municipalities. *Mathematical Problems in Engineering*.
- Tang, Y., W. Zhao and C. Gu. 2017. Urbanization and rural development in the Beijing–Tianjin–Hebei metropolitan region: Coupling-degree model. *Journal of Urban Planning and Development* 143: 04016028.
- Tang, Z. 2015. An integrated approach to evaluating the coupling coordination between tourism and the environment. *Tourism Management* 46: 11–19.
- Takeuchi, H. 2013. Survival strategies of township governments in rural China: From predatory taxation to land trade. *Journal of Contemporary China* 22: 755–772.
- Tesfaye, M. A., A. Bravo-Oviedo, F. Bravo, B. Kidane, K. Bekele and D. Sertse. 2015. Selection of tree species and soil management for simultaneous fuelwood production and soil rehabilitation in the Ethiopian central highlands. *Land Degradation & Development* 26: 665–679.
- Tian, L., B. Ge and Y. Li. 2017. Impacts of state-led and bottom-up urbanization on land use change in the peri-urban areas of Shanghai: Planned growth or uncontrolled sprawl? *Cities* 60: 476–486.
- Tyner, A. and Y. Ren. 2016. The Hukou system, rural institutions, and migrant integration in China. *Journal of East Asian Studies* 16: 331–348.
- Wang, S., H. Ma and Y. Zhao. 2014. Exploring the relationship between urbanization and the eco-environment—A case study of Beijing–Tianjin–Hebei region. *Ecological Indicators* 45: 171–183.
- Wang, X., E. C. Hui and J. Sun. 2017a. Population migration, urbanization and housing prices: Evidence from the cities in China. *Habitat International* 66: 49–56.
- Wang, X., S. Zhao and J. Chen. 2017b. Population agglomeration, economic growth and environmental degradation: Evidence from China's 290 cities during 2003–2012. *Applied Ecology and Environmental Research* 15: 1511–1525.
- Wu, K. and H. Zhang. 2012. Land use dynamics, built-up land expansion patterns, and driving forces analysis of the fast-growing Hangzhou metropolitan area, eastern China (1978–2008). *Applied Geography* 34: 137–145.
- Xiao, R., S. Su, Z. Zhang, J. Qi, D. Jiang and J. Wu. 2013. Dynamics of soil sealing and soil landscape patterns under rapid urbanization. *Catena* 109: 1–12.
- Xie, C., X. Huang, H. Mu and W. Yin. 2017. Impacts of land-use changes on the lakes across the Yangtze floodplain in China. *Environmental Science & Technology* 51: 3669–3677.
- Xu, H., X. Wang and G. Xiao. 2000. A remote sensing and GIS integrated study on urbanization with its impact on arable lands: Fuqing City, Fujian Province, China. *Land Degradation & Development* 11: 301–314.
- Xu, X. and X. Min. 2013. Quantifying spatiotemporal patterns of urban expansion in China using remote sensing data. *Cities* 35: 104–113.
- Zhang, J. 2011. Interjurisdictional competition for FDI: The case of China's development zone fever? *Regional Science and Urban Economics* 41: 145–159.
- Zhou, Z. 2016. Overreaction to policy changes in the housing market: Evidence from Shanghai. *Regional Science and Urban Economics* 58: 26–41.

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



IN THE AERIAL IMAGERY WORLD THIS IS SANTA'S WORKSHOP

FREE TRIAL OFFER

Check out our massive cloud-based library of 2-3" imagery through our viewing/measuring tools, API or our ArcGIS plug in. Take advantage now.