

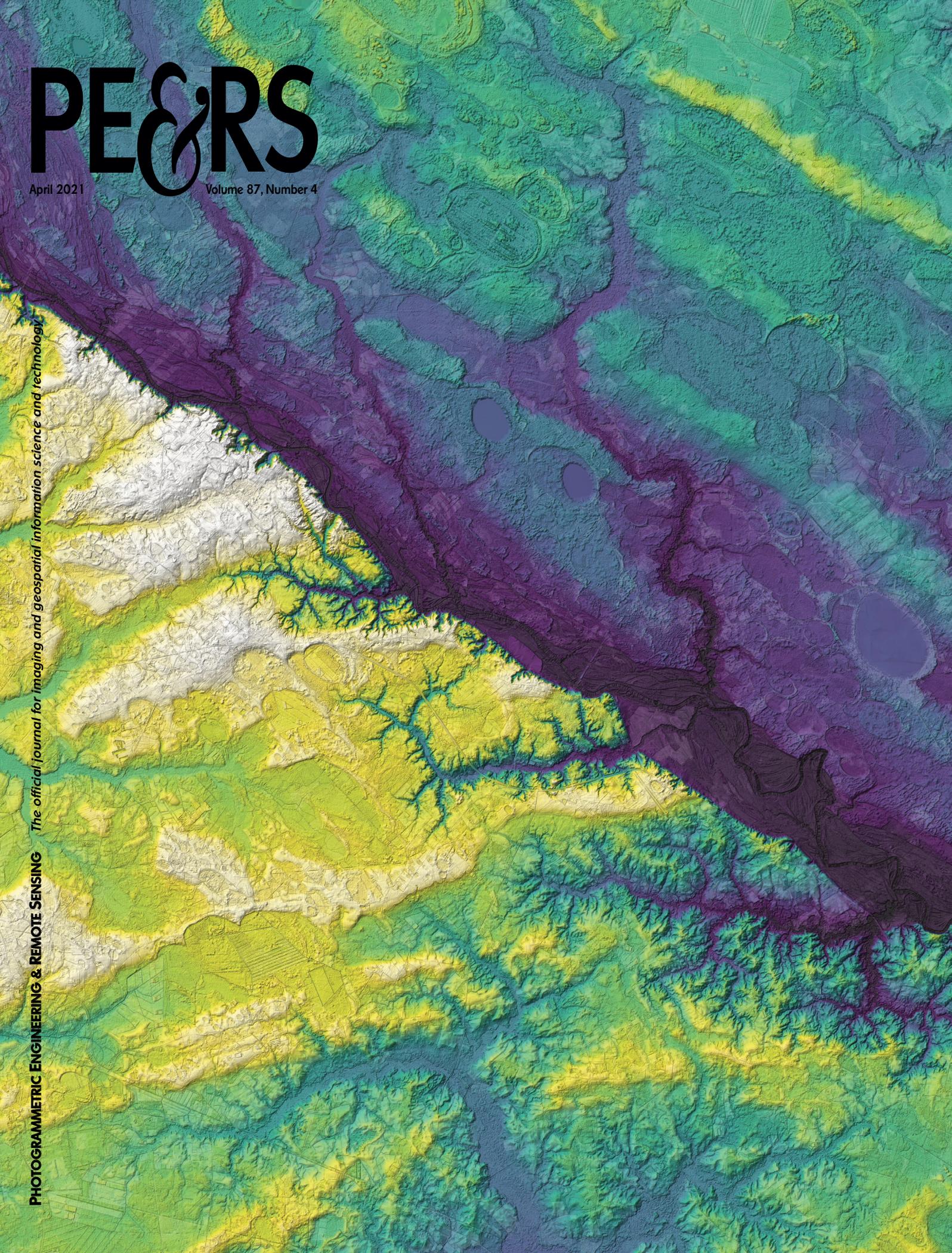
# PE&RS

April 2021

Volume 87, Number 4

*The official journal for imaging and geospatial information science and technology*

**PHOTOGRAMMETRIC ENGINEERING & REMOTE SENSING**





# XXIV ISPRS Congress

International Society for Photogrammetry and Remote Sensing

## 4 – 10 July 2021, Nice, France



**COMMISSION I**  
Sensor Systems

GERMANY / BRAZIL

**COMMISSION II**  
Photogrammetry

ITALY / JAPAN

**COMMISSION III**  
Remote Sensing

CHINA / CANADA

**COMMISSION IV**  
Spatial Information Science

THE NETHERLANDS / CANADA

**COMMISSION V**  
Education and Outreach

INDIA

[www.isprs2020-nice.com](http://www.isprs2020-nice.com)

### Keynote speakers

 <b>CHEN JUN</b> Chief Scientist NGCC	 <b>YVES UBELMANN</b> CEO - ICONEM	 <b>WILLEMIN SIMON VAN LEEUWEN</b> Founder and CEO GeoFort	 <b>RAQUEL URTASUN</b> Head of Uber ATG	 <b>MICHAEL TONON</b> AIRBUS DEFENCE & SPACE	 <b>DAVIDE SCARAMUZZA</b> Professor University of Zurich & ETH Zurich	 <b>PIERRE ALLIEZ</b> Head of Science at Inria Sophia Antipolis
---	--	--	---	--	---	---

### Executive committee

 <b>Nicolas Paparoditis</b> ENSG-IGN-Univ. Gustave Eiffel Congress Director	 <b>Laurent Polidori</b> CESBIO Congress Deputy Director	 <b>Tania Landes</b> INSA Congress Media Chair	 <b>Clément Mallet</b> IGN Congress Program Chair	 <b>Florent Lafarge</b> INRIA Congress Program co-Chair	 <b>Pierre Grussenmeyer</b> INSA Congress Industry Chair	 <b>Eric Labergerie</b> CNAM Congress Financial Chair
--	---	---	--	--	---	--

Platinum sponsor  
**AIRBUS**

Gold sponsors  
**ThalesAlenia Space** **esri**  
THE SCIENCE OF WHERE™

Silver sponsors  
**HEXAGON**

Bronze sponsors

Institutional partners, supporting graduate schools and associations

## ANNOUNCEMENTS

NV5 Geospatial, powered by **Quantum Spatial**, today launched INSITE, a comprehensive cloud-based platform that enables users to more efficiently and cost-effectively manage their geospatial data, from acquisition to delivery. With applications designed for electric utilities and telecommunications, as well as federal, state and local governments, INSITE provides tools to support the entire geospatial data lifecycle, including project tracking, data collection and delivery, quality control, on-demand reporting, analytics and enterprise integration.

“Location-based data is critical for business and governments when it comes to managing risk, lowering costs, forecasting work and maintaining compliance. Yet there are numerous technical and organizational barriers that prevent them from optimizing use of this information,” said Mark Abatto, president and COO of NV5 Geospatial. “INSITE was purpose-built to break down these barriers, creating a single source of truth for all geospatial data within an organization. Now, with INSITE, distributed teams can collaborate more effectively, visualize data with ease and gain real-time insights through powerful analytics.”

INSITE enables users to import, search, analyze, manage, integrate and export all types of geospatial data and multimedia. As a cloud-based platform, INSITE improves speed and efficiency, minimizes storage expenses and supports greater collaboration by eliminating the use of standalone hard drives or a dedicated computer.

The INSITE platform is the foundation for three primary categories of use. All users can benefit from INSITE Lifecycle, which provides visibility into project status through a web-based map with real-time tracking of data from acquisition through processing and delivery. INSITE Core enables data visualization and interaction, including access to aerial imagery, lidar data, GIS layers, and more. Lastly, INSITE provides the basis for Industry-specific applications, including:

- INSITE Facility, provides configurable asset management, including planning, execution, and tracking for inspections and maintenance on any asset and facility.
- INSITE Pole Manager, which enables electric utilities to take advantage of automated pole modeling analytics that support joint use, clearance, and integrity analysis programs

INSITE turbocharges use of geospatial data through integration with enterprise resource and asset management systems, such as Esri, IBM, Oracle, SAP, SPIDAcad and others. Compatible with all major operating systems, INSITE is accessible via any Internet-connected device. To learn more, visit [insitenv5g.com/insite](http://insitenv5g.com/insite).

**GeoCue Group** announces the release of their latest addition to their 3D Imaging Systems (3DIS<sup>®</sup>) series, the True View<sup>®</sup> 515. This compact, affordable 3D Imaging System has significantly better sensitivity compared to other sensors in its class, resulting in remarkable analytic data. True View 515 will redefine lidar wire extraction and dense vegetation penetration

applications for all-purpose sensors. GeoCue is excited to bring this state-of-the-art 3D imaging system to the market.

GeoCue changed the drone lidar/Imaging market in 2019 with the creation of a 3D Imaging Sensor/System approach to drone mapping. Unlike conventional drone lidar systems with tacked on cameras, a True View 3DIS<sup>®</sup> couples a laser scanner with tightly integrated and calibrated photogrammetric cameras. The included post-processing software generates RGB colorized point clouds from the source images. Unlike the flat colorization from ortho images seen in other systems, the True View 3DIS<sup>®</sup> renders 3D colorization for spectacular, high accuracy data sets.

GeoCue's True View 515 is unlike any sensor in its class. The True View 515 series is an all-purpose grade 3DIS<sup>®</sup> designed for small Unmanned Aerial Systems. The system includes full post-processing software that generates a ray-traced 3D colorized point cloud and geocoded images. GeoCue's True View 3DIS<sup>®</sup> product series is an inclusive hardware and software offering providing users with a complete acquisition to deliverables workflow.

“Our True View 410 3DIS<sup>®</sup> has created a new standard for fused lidar and imagery in an affordable, easy to use package and our RIEGL miniVUX-based systems are really the best available drone sensors for survey grade projects” said Lewis Graham, President/CTO of GeoCue. “However, in discussions with our customers, we kept hearing about the gap between these two products; clients wanted to collect transmission wires and see through heavy vegetation and would accept accuracies/precision slightly below survey grade. After testing many high beam count sensors, we were able to achieve spectacular results with the True View 515 configuration. There is simply no drone lidar/Imaging system under \$100,000 that will produce data comparable to the True View 515.”

True View 515 3DIS<sup>®</sup> is equipped with Hesai's PandarXT-32 laser scanner integrated with dual GeoCue photogrammetric cameras and Applanix APX-15 positioning system. The True View product line gives surveyors and mappers the ability to collect and deliver high quality data. These deliverables are generated using workflows and tools within the included True View EVO post-processing software.

Examples of derived products include break line enforced models, profiles and cross sections, topographic contours, volumetric analysis, classified ground model and more. Every True View 3DIS<sup>®</sup> includes all the software needed to flow from raw collected data to product deliverables; True View EVO, Applanix POSPac and True View Reckon post-deployment data management portal.

GeoCue's True View 3DIS<sup>®</sup> are designed to create point cloud data that have been colorized with Red-Green-Blue (RGB) camera data via a rigorous point by point tracing algorithm. One of the biggest advantages of a True View 3DIS<sup>™</sup> is the speed with which these data can be produced; a 15-minute data collect can be processed to a colorized point cloud in about 10 minutes! To learn more, visit [www.geocue.com](http://www.geocue.com).

## APPLICATIONS

TCarta Marine, a global provider of space-based hydrospace solutions, has been awarded a Small Business Innovation Research (SBIR) Phase II grant from the National Oceanic and Atmospheric Administration (NOAA). The research focuses on enhancing Satellite Derived Bathymetry technology for application in the coastal waters of Alaska.

Satellite Derived Bathymetry (SDB) extracts water depth measurements from multispectral satellite imagery using advanced physics-based algorithms. Traditionally, this technology has yielded the most accurate results in clear, calm waters; however, TCarta has recently taken the lead in expanding SDB applications to more challenging marine environments worldwide.

“With the NOAA SBIR grant, our objective is to explore what SDB can achieve in Alaska’s coastal waters with the expectation of applying our findings to the greater Arctic region,” said TCarta President Kyle Goodrich. “This research will benefit hydrographic surveying, coastal zone management, infrastructure development, and other coastal marine projects at high latitudes.”

Begun in 2020, the NOAA SBIR Phase I research expanded the potential sources of satellite imagery for the Alaskan coast. Winter darkness, floating sea ice, turbulent spring runoff, plankton blooms, and unpredictable tidal variations make it difficult to capture images with the quality necessary for SDB processing. Just a small window of acquisition time exists during the summer, and it varies for different areas along Alaska’s 32,000 miles of coastline.

“By analyzing data captured with several types of satellite sensors, including multispectral, hyperspectral, green laser, and synthetic aperture radar, we are determining which images and methodologies – or combinations – produce the optimal characteristics for SDB,” said Goodrich. “This is helping us devise a tasking strategy to capture Alaska imagery at the right time with the right sensors.”

An important Phase I result was the development of automated tools to assess water clarity in daily satellite images to monitor water clarity for suitable conditions for SDB image tasking or planning airborne lidar surveying, Goodrich added.

In Phase II of the NOAA program, TCarta has shifted focus to SDB data processing and hydrographic analysis by developing new ways to improve confidence in the seafloor depth measurements derived from Alaskan imagery. The Colorado firm is creating workflows involving Artificial Intelligence to identify and minimize errors in SDB measurements, predict uncertainty and to integrate established hydrographic standards, thereby enhancing interoperability of SDB with other technologies.

“By the start of summer 2022 collection season, we will have developed a multi-satellite coastal surveying capability and a collection strategy for maximization of these technologies in Alaskan and Arctic waters,” said Goodrich. For more information about TCarta, visit [www.tcarta.com](http://www.tcarta.com).

The Juneau Icefield Research Program (JIRP) calculates that thinning of Alaska’s Taku Glacier has increased from an average rate of a half meter to two meters per year over the past two decades. Annual mapping by JIRP reveals the glacier’s thickness has increased and decreased from one year to the

next, likely due to snow accumulation variability, but the overall current trend shows an annual net loss of ice.

“Taku is losing enough meltwater every day to fill an NFL stadium,” said Seth Campbell, JIRP Director of Academics & Research.

JIRP monitors the complex kinematics and mass balance of the Juneau Icefield – that is, they study annual changes to ice velocity, snow accumulation, and surface melting – to estimate if the glacier is advancing or retreating over time. One way they accomplish this task is by mapping yearly GPS field points in the Grapher and Surfer modeling packages from Golden Software, a developer of affordable 2D and 3D scientific software based in Golden, Colo.

“JIRP’s study of the many processes impacting the Juneau Icefield demonstrates the value of applying advanced visualization to gain insights from complex data sets,” said Golden Software CEO Blakelee Mills.

Accurate assessment of the Taku melt rate is vital to the study of climate change because, at more than 800 square kilometers, the glacier is the largest in the massive Juneau Icefield. Straddling the Alaska-Canada border, the receding icefield plays multiple important roles in local ecosystems. For the residents of British Columbia, it is a reliable supply of freshwater, but for the Gulf of Alaska, increasing glacier meltwater can have potentially negative impacts the marine ecosystem and valuable fisheries.

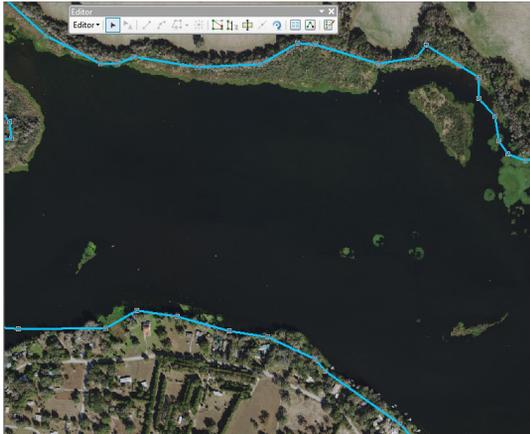
JIRP research dates to 1946, but the introduction of GPS contributed significantly to annual summer fieldwork beginning in 1993. Volunteers capture more than 1,000 GPS measurements at designated transect locations on the icefield each year to record glacial velocity and surface elevation changes. These values reveal whether the glacier is growing or shrinking due to changes in precipitation or melt rates or from changing rate of movement.

“We have seen increases in thickness after particularly snowy winters,” said Scott McGee, JIRP Geomatics Program Lead. “But less snow accumulation and more surface melt seem to be the primary drivers of change now.”

JIRP researchers rely on Golden Software’s Grapher scientific graphing package to plot GPS “Z” elevation values across transects in 2D to generate thickness profiles. They often overlay multiple years’ worth of field data to illustrate change over time. The scientists also input GPS field points for multiple transects from multiple years into the Surfer 3D surface mapping package to gain a sense of overall glacier volume change.

Together, these ice elevations, velocities, and volume change calculations give researchers a better understanding of external forces acting upon the Juneau Icefield and what to expect in the future. Read the full case study at <https://www.goldensoftware.com/success-stories/tracking-glacier-thickness-changes-with-grapher-and-surfer>.

Golden Software’s Grapher and Surfer packages are used by more than 100,000 scientists and engineers worldwide, many involved in oil & gas exploration, environmental consulting, mining, engineering, and geospatial projects. The software has been relied upon for more than 35 years by users in numerous disciplines to easily visualize and interpret complex data sets. For additional information, visit [www.GoldenSoftware.com](http://www.GoldenSoftware.com).



## 225 GIS Tips & Tricks— Using GIS to Hunt for Easter Eggs

### COLUMNS

- 227 Mapping Matters
- 229 Book Review—*GIS: An Introduction to Mapping Technologies, First Edition*
- 231 Grids and Datums Update  
This month we look at the Co-operative Republic of Guyana
- 233 Signatures  
The Column of the Student Advisory Council

### ANNOUNCEMENTS

- 228 New ASPRS Members  
Join us in welcoming our newest members to ASPRS.
- 234 Headquarters News
  - 234 ASPRS Announces Achievement Awards
  - 236 Welcome Shalini Gogawale!
  - 236 The ASPRS Mid-South Region Proudly Announces the Formation of the Auburn University Student Chapter
  - 236 ASPRS Announces Henry John “Hank” Theiss as the 2021 Recipient of the Photogrammetric (Fairchild) Award
- 272 Call for Submissions—*PE&RS* Special Issue on Remote Sensing Monitoring for Urban Environment

### DEPARTMENTS

- 221 Industry News
- 236 Ad Index
- 248 In-Press *PE&RS* Articles
- 294 ASPRS Sustaining Members

#### 237 A Digital Terrain Modeling Method in Urban Areas by the ICESat-2 (Generating precise terrain surface profiles from photon-counting technology)

*Nahed Osama, Bisheng Yang, Yue Ma, and Mohamed Freeshah*

The ICE, Cloud and land Elevation Satellite-2 (ICESat-2) can provide new measurements of the Earth’s elevations through photon-counting technology. Most research has focused on extracting the ground and the canopy photons in vegetated areas. Yet the extraction of the ground photons from urban areas, where the vegetation is mixed with artificial constructions, has not been fully investigated. This article proposes a new method to estimate the ground surface elevations in urban areas.

#### 249 A Novel Class-Specific Object-Based Method for Urban Change Detection Using High-Resolution Remote Sensing Imagery

*Ting Bai, Kaimin Sun, Wenzhuo Li, Deren Li, Yepi Chen, and Haigang Sui*

A single-scale object-based change-detection classifier can distinguish only global changes in land cover, not the more granular and local changes in urban areas. To overcome this issue, a novel class-specific object-based change-detection method is proposed.

#### 263 Discovering Potential Illegal Construction Within Building Roofs from UAV Images Using Semantic Segmentation and Object-Based Change Detection

*Yang Liu, Yujie Sun, Shikang Tao, Min Wang, Qian Shen, and Jiru Huang*

A novel potential illegal construction (PIC) detection method by bitemporal unmanned aerial vehicle (UAV) image comparison (change detection) within building roof areas is proposed.

#### 273 Error Analysis and Optimization of a Sky Full-Polarization Imaging Detection System

*Yongtai Chen, William C. Tang, Jinkui Chu, Ran Zhang, and Song Li*

An accurate sky polarization field map is a prerequisite for polarization navigation applications. In this article, a detector for sky full-polarization imaging detection is described, the major error-influencing factors (MEIFS) are obtained, and the error propagation is modeled and analyzed.

#### 283 Parsing of Urban Facades from 3D Point Clouds Based on a Novel Multi-View Domain

*Wei Wang, Yuan Xu, Yingchao Ren, and Gang Wang*

Recently, performance improvement in facade parsing from 3D point clouds has been brought about by designing more complex network structures, which cost huge computing resources and do not take full advantage of prior knowledge of facade structure. Instead, from the perspective of data distribution, we construct a new hierarchical mesh multi-view data domain based on the characteristics of facade objects to achieve fusion of deep-learning models and prior knowledge, thereby significantly improving segmentation accuracy.

#### 295 Scene Classification of Remotely Sensed Images via Densely Connected Convolutional Neural Networks and an Ensemble Classifier

*Qimin Cheng, Yuan Xu, Peng Fu, Jinling Li, Wei Wang, and Yingchao Ren*

Deep learning techniques, especially convolutional neural networks, have boosted performance in analyzing and understanding remotely sensed images to a great extent. However, existing scene-classification methods generally neglect local and spatial information that is vital to scene classification of remotely sensed images. In this study, a method of scene classification for remotely sensed images based on pretrained densely connected convolutional neural networks combined with an ensemble classifier is proposed to tackle the underutilization of local and spatial information for image classification

**See the Cover Description on Page 224**

# COVER DESCRIPTION



During the last ice age, the Atlantic coastal plain of the southeastern U.S. was a cold, dry, and windy place. With patches of permafrost stretching as far south as Georgia, the seasonal freezing and thawing of sandy soils left landscapes littered with thermokarst lakes and scattered sand dunes and ridges—much like those now found in parts of Alaska and the Arctic.

If you know where to look in the Carolinas, Georgia, Virginia, Maryland, and Delaware, you can still find abundant relicts from this cooler period. At ground level, clusters of low-relief oval depressions called Carolina bays can look as ordinary as a lake, bog, or farm field. Yet from above—from an airplane or satellite—the distinctive features stand out.

The image at the top of the page shows a collection of Carolina bays near Bladen Lakes State Forest in North Carolina. The natural-color image was acquired by the Operational Land Imager (OLI) on Landsat 8; the data were then draped over a digital elevation model to emphasize the terrain.

Many of the Carolina bays in this area have elliptical shapes oriented in a south-east-northwest direction. This orientation is a product of the prevailing winds when the lakes were forming, explained Christopher Swezey of the U.S. Geological Survey. “The dominant wind direction in this part of North Carolina was from the west, and that wind set up gyres in the lakes that eroded shorelines on the northwest and southeast margins of the lake.”

According to Swezey, most Carolina bays likely formed episodically during the last glacial maximum and the beginning stages of deglaciation, a period that spanned 40,000 to 11,000 years ago. While other geologists have speculated about the Carolina bays possibly being products of meteor impacts, Swezey has amassed evidence that they are relict thermokarst lakes.

An early study of the features estimated there are as many as 500,000 Carolina bays in the Atlantic coastal plain. “We now know that was a really conservative estimate,” said Swezey. “With the availability of modern lidar data, we are seeing many more of them than were previously recognized.”

The broad distribution of the Carolina bays suggests that permafrost extended for several hundred kilometers south of the ice sheet. “This is noteworthy because many geologists who studied glaciation in the past have suggested that frozen ground extended only as far south as northern Virginia during the last ice age,” said Swezey.

For more information and to see additional imagery, visit <https://landsat.visibleearth.nasa.gov/view.php?id=147904>.

NASA Earth Observatory images by Joshua Stevens, using Landsat data from the U.S. Geological Survey and topographic data from North Carolina State University. Story by Adam Voiland.



## PHOTOGRAMMETRIC ENGINEERING & REMOTE SENSING

JOURNAL STAFF

Publisher ASPRS

Editor-In-Chief Alper Yilmaz

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](http://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 2021 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 rate of the e-Subscription (digital) Site License Only for USA and Non-USA is \$1000.00 USD. e-Subscription (digital) Site License Only for Canada\* is \$1049.00 USD. e-Subscription (digital) Plus Print for USA is \$1365.00 USD. e-Subscription (digital) Plus Print for Canada\* is \$1424.00 USD. e-Subscription (digital) Plus Print for Non-USA is \$1395.00 USD. Printed-Subscription Only for USA is \$1065.00 USD. Printed-Subscription Only for Canada\* is \$1124.00 USD. Printed-Subscription Only for Non-USA is \$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, \$65.00 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 \$150.00 USD; for Student members \$50.00 USD for USA and Canada \$53.00 USD; \$60.00 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 2021.** 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.

# GIS Tips & Tricks

Al Karlin, Ph.D. CMS-L, GISP, Keith Patterson, PSM, SP, GISP, Carly Bradshaw, Savannah Carter and Todd Waldorf

## Using GIS to Hunt for Easter Eggs

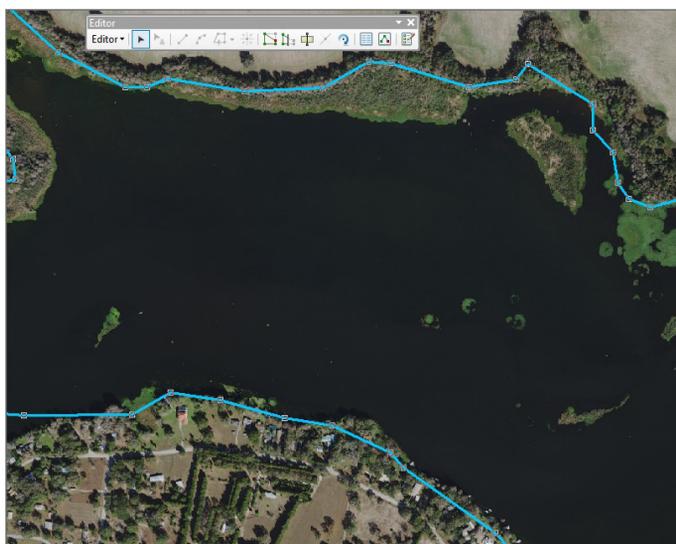
While writing GIS software is a big business and no laughing (or trivial) matter, ever since the early days of computer software programs, the authors have displayed their sense of humor, and creativity, by placing little “hidden features or Easter eggs” in their coding. Sometimes these hidden

features would display the coders names when a special key sequence or click-pattern was detected. Other times, hidden features, likened to Easter eggs, would be revealed with key combinations. In this month’s column, we highlight five GIS Tips & Tricks that are easy to miss.

### Tip #1:

(From Carly Bradshaw)

Carly does a lot of line-vector editing and discovered that when in an active editing session in ArcGIS, if you hover over a line and toggle the “V” key on the keypad, all of the vertices appear/disappear, as below:

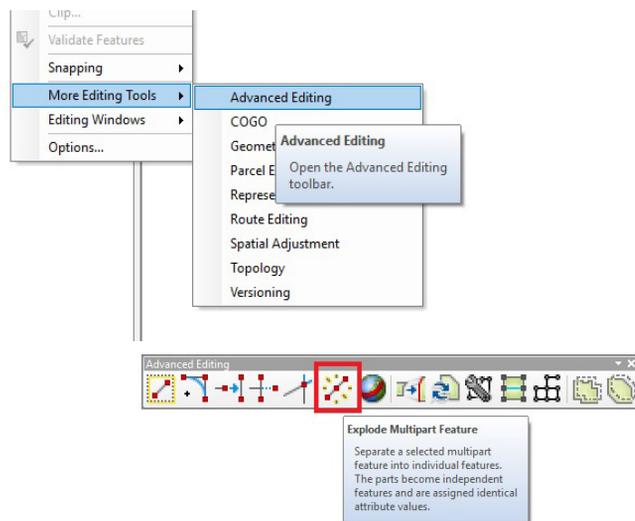


Pressing/releasing the V-key toggles the vertices on and off but only when hovering the cursor near the line (or polygon) being edited.

### Tip #2:

(From Savannah Carter)

Like Carly, Savannah performs a lot of vector editing, and found that she often forgets about “Exploding Multipart Feature”. As with most tools, there are multiple ways to enable the “Explode Multipart Feature”. The tool is available in the ArcToolbox | Data Management, and on the Advanced Editing toolbar as:



Of course, the tool is found in the Advanced Editing Customize menu and can be added to any toolbar.

Photogrammetric Engineering & Remote Sensing  
Vol. 87, No. 4, April 2021, pp. 225–226.  
0099-1112/21/225–226

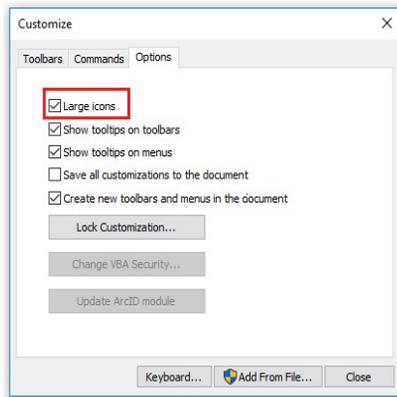
© 2021 American Society for Photogrammetry  
and Remote Sensing  
doi: 10.14358/PERS.87.4.225

# Tip #3

(from Al Karlin)

While I am at it, you may have noticed that the size of the icons on the Advanced Editing Toolbar above is larger than the standard display. That is not a result of me enlarging a picture with some digital Paint program. For those of us with aging eyes or very high-resolution computer displays, there is an option to increase the size of ALL of the icons on the Esri toolbars. While this option is not exactly hidden, it is not very obvious either.

To increase the size of the icons, from the Main Toolbar use the Customize | Customize Mode (or double-click anywhere on the gray space on the interface) to open the “Customize” dialog. Then click on the Options tab and select “Large Icons.”



# Tip #4

(from Todd Waldorf)

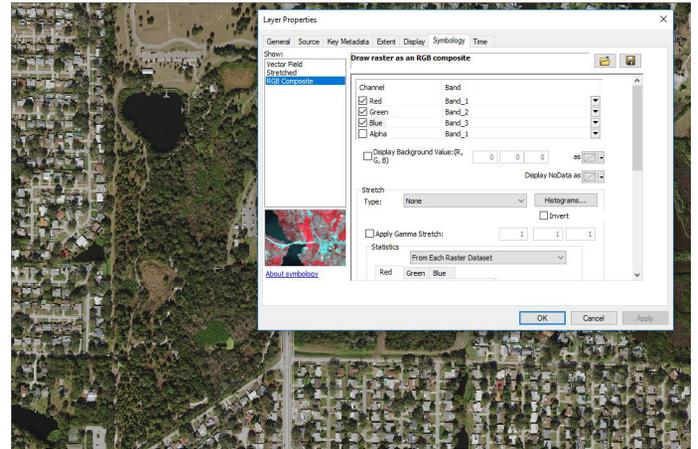
Todd works with many GIS software packages at Dewberry and has contributed some other Tips & Tricks to this column. While editing vector files, Todd finds that zooming in and out with the mouse wheel is sometimes too coarse for his needs. Todd discovered that holding the <CTRL> key down while wheeling will provide finer control over the zoom increment.

# Tip #5

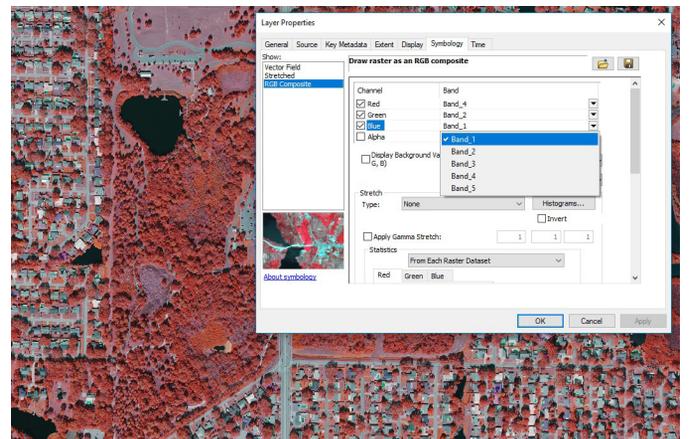
(from Keith Patterson)

Keith spends a lot of time reviewing ortho-imagery and associated products as a final quality check prior to delivery to clients. While this tip is not technically an “Easter egg”, Keith found that when reviewing 4-band multichannel imagery, it is really easy to “switch” between bands with the Layer Properties menu.

In the Esri ArcGIS Software, the natural color image is rendered by assigning Red to Band 1, Green to Band 2, and Blue to Band 3, as below:



To render the multi-band image as a false-color Infrared image, use the dropdown arrows in the Layers | Symbology tab and assign the infrared, usually, Band 4 to Red, leave Green assigned to Band 2, and assign red to Band 3 as below:



Of course, click “OK” to apply the new band/color assignments.

Easter eggs can be found in all software packages. Please feel free to share yours with us. Send your questions, comments, and tips to GISTT@ASPRS.org.

**Al Karlin, Ph.D., CMS-L, GISP, Keith Patterson, PSM, SP, GISP, Savannah Carter, Carly Bradshaw, and Todd Waldorf** are with Dewberry’s Geospatial and Technology Services group in Tampa, FL. As a senior geospatial scientist, Al works with all aspects of lidar, remote sensing, photogrammetry, and GIS-related projects. Keith, a Florida-licensed Professional Surveyor and Mapper, manages photogrammetry, lidar, and GIS mapping projects nationwide. Carly is a Geospatial Analyst who works on lidar and breakline editing. Savannah, also a Geospatial Analyst, specializes in topographic and bathymetric lidar projects. Todd, a Geospatial Analyst too, works as a photo-interpreter and lidar editor (and is incredibly good at finding Easter eggs before younger GIS analysts).



# MAPPING MATTERS

## YOUR QUESTIONS ANSWERED

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

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

### QUESTION:

**Question:** Do frame-based remote sensing systems, such as metric cameras, rely significantly on accurate inertial measurement unit (IMU) data with OPK (omega, phi, cappa) parameters for the creation of high-accuracy geospatial products? Or is position (XYZ) alone more important, as long as aerotriangulation can solve frame rotation during post-processing?

Nathan Eick, Flight Operations Manager,  
Aerial Services, Inc. (ASI)

**“Experience has shown us that while direct orientation sometimes achieves project accuracy, it fails to do so at other times.”**

**Dr. Abdullah:** My answer to your question Nathan is no, an aerial metric camera does not rely significantly on IMU orientation, as long as you are performing aerial triangulation on the imagery. For processing imagery, IMU has a significant advantage if you are using the direct orientation concept to bypass the process of aerial triangulation. However, you cannot always rely on the quality of the solution from the direct orientation for accurate mapping. Experience has shown us that while direct orientation sometimes achieves project accuracy, it fails to do so at other

**“The concept of direct orientation hurts the most when you use it to set up stereo models for compilation”**

times. The concept of direct orientation hurts the most when you use it to set up stereo models for compilation. Some of these stereo pairs come clean and are free of parallax, while many other stereo models show severe parallax that hinders the completion of the map. That then begs the question, “What if I use high-end IMU that is very accurate in determining sensor orientation angles?” High-end IMU is sometimes more accurate than aerial triangulation in determining sensor orientation. However, modern imaging

**“During aircraft takeoff and landing, not to mention soaring into the upper atmosphere, there are always mechanical and thermal stresses that cause vibration on and between these system components”**

processes rely on multiple elements, including the camera, GPS receiver, IMU, camera mount, laptop, etc. The presence of parallax in some stereo pairs could be caused by a source other than IMU. When we put IMU on the digital camera, we are bolting the IMU hardware, i.e. box, to the camera body or the assembly that carries the lens. During aircraft takeoff and landing, not to mention soaring into the upper atmosphere, there are always mechanical and thermal stresses that cause vibration on and between these system components. While a good IMU measures accurate orientation angles of its body frame, these mechanical stresses may cause disruption in accurately transforming the orientation of the camera body and vice versa. Having accurate, IMU-derived orientation angles is always beneficial to the

**“While a good IMU measures accurate orientation angles of its body frame, these mechanical stresses may cause disruption in accurately transforming the orientation of the camera body and vice versa”**

aerial triangulation solution as it can reduce the number of ground control points and it may accelerate the image matching process. However, the presence of IMU in aerial triangulation is not as important as utilizing accurate GPS positions. Having an accurate GPS position for each frame

Photogrammetric Engineering & Remote Sensing  
Vol. 87, No. 4, April 2021, pp. 227–228.  
0099-1112/21/227–228

© 2021 American Society for Photogrammetry  
and Remote Sensing  
doi: 10.14358/PERS.87.4.227

**“Utilizing accurate GPS positions and a few ground control points are all we need to perform successful aerial triangulation”**

is crucial to processing imagery through aerial triangulation. This replaces the need to survey a tremendous amount of ground control points previously needed for the old practices of aerial triangulation prior to the invention of GPS. IMU can be handy for single flight-line corridor projects because it can reduce the amount of ground control points needed for processing the imagery. However, using high-end IMU for metric, large-format cameras can cost hundreds of thousands of dollars. This is why many of us in the industry stopped adding IMU to the camera after we experienced the problem of parallax with direct orientation. Utilizing accurate GPS positions and a few ground control points are all we need to perform successful aerial triangulation.

*\*\*Dr. Abdullah is Vice President and Chief Scientist at Woolpert, Inc. He is also adjunct professor at Penn State and the University of Maryland Baltimore County. Dr. Abdullah is ASPRS fellow and the recipient of the ASPRS Life Time Achievement Award and the Fairchild Photogrammetric Award.*

The contents of this column reflect the views of the author, who is responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the American Society for Photogrammetry and Remote Sensing, Woolpert, Inc., NOAA Hydrographic Services Review Panel (HSRP), Penn State, and/or the University of Maryland Baltimore County.

## NEW ASPRS MEMBERS

ASPRS would like to welcome the following new members!

### At Large

Mubashir Ali  
Prawin Balasubramaniam  
Mohammadreza Bayanlou  
Pranavan Bupathy  
SP Dhanabalan  
Abhinav Galodha  
Kalingga Titon Nur Ihsan  
Yogendra Karna  
Mehdi Khoshboresh-Masouleh  
Eslam M. Muhammed  
Jokotola Toyeybat Omidiji  
Mike Petrut  
Surachet Srinara  
Aswin Sunil  
Sagar Taneja  
Prateek Tripathi  
Ravi Verma

### Cascadia

Sara Mei Ling Goldstein  
Brooke Hunter  
Josh Wiejaczka

### Eastern Great Lakes

Casey Lawrence Blake

### Florida

Connor Bass  
Fred Hartless, GISP  
Mason Jordan Knapp  
Troy Lane Underhill

### Heartland

Katie Buchok  
Kenneth Ekpeteru  
Morgan Lu  
Douglas MacFarland  
Wenjun Yang

### Mid South

Gregory Alliger  
Jason Combs  
Sukanya Dasgupta  
Jacob Levi Gresham  
Bilal Gregory Jones  
Keiyonna LaChelle Lighten-Solomon  
Kyle Manderson, CP, CMS  
John Segars  
Xiaoqing Shen  
Megha Shrestha

Kama Svoboda

Esra Tekdal Tekdal Yilmaz, PhD  
Jonathan Tidwell  
Samuel Wyatt  
Kristin Youngquist

### North Atlantic

MD Amit Hasan  
Michael S. Loose, CP  
Joshua Persson, CP

### Northeastern

William M. Stiteler, CMS, PhD  
Cora Van Hazinga

### Pacific Southwest

Steve Hannig  
Kara Lee  
Alejandro Vasquez  
Matthew Winter

### Potomac

Gabriel Antonio Abreu-Vigil  
Arif Albayrak  
Steven R. Douglas  
Christopher D. Gard  
Yvonne Harding  
Ella Kasten  
Laura Mills  
Hope S. Morgan, PLS  
Casey Thomas O'Heran  
Carl David Stearns  
Brian Sweeney  
Neri G. Terry, Jr.  
Marlin R. Zook, CP

### Rocky Mountain

Abigail Gettinger  
James Haynes  
Christopher M. McGinty  
Brooks W. Mitchell  
Christopher Payne

### Western Great Lakes

Suzanne Liebergen  
Stephen J. Martinez, CP  
Kyle Mullen  
Bryan Tolcser  
Gabriel David Wier  
Cyril Oluyomi Norwood Wilson, PhD  
Luke Zaruba

FOR MORE INFORMATION ON ASPRS MEMBERSHIP,  
VISIT [HTTP://WWW.ASPRS.ORG/JOIN-NOW](http://www.asprs.org/join-now)

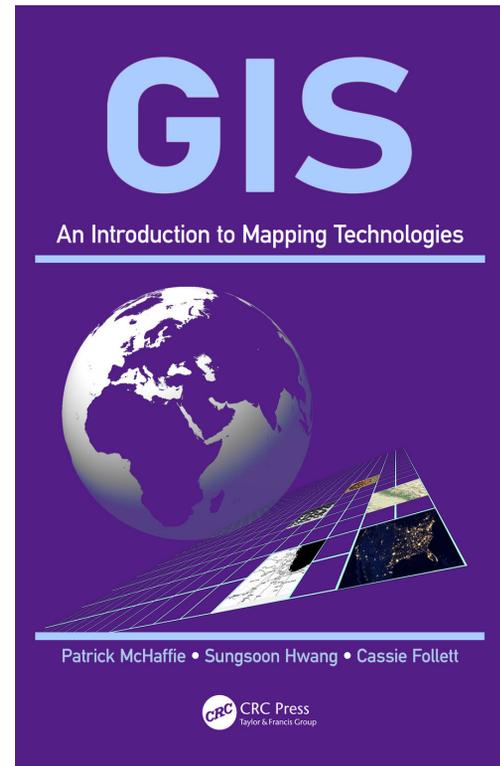
**Your Path To Success In The Geospatial Community**

This book is a wide-ranging and wide-open introduction to geographic information systems technology appropriate for undergraduate and graduate courses or for professionals interested in getting started with this technology or broadening their knowledge of the field.

This book is organized into twelve chapters and covers a wide variety of material that quickly brings a student up to speed on the most important GIS ideas and concepts. Most chapters include exercises in a section at the end that illustrates the concepts covered. The chapters were written by different authors, which is evident in some organizational and style inconsistencies, but this is more than made up for by the amount and depth of included content. The chapters are not all purely technical content, the text is bookended by Chapters 1 and 12, which provide context for students and professionals about the why and how of working with geospatial technology.

Chapter 1—*Making Sense of Geotechnology* provides a vision of the purpose of GIS technology, the role of the GIS professional, and the purpose of the content of this book. Chapter 2—*Georeferencing* covers the basics of coordinate systems and how they are applied to geospatial data. Chapter 3—*Getting the World into Your GIS* discusses how data are acquired, including digitizing, GNSS, and remotely sensed data. Chapter 4—*Geospatial Analysis: Introduction* introduces the topic through operations of selection, proximity, and interpolation. Chapter 5—*Thematic Mapping* introduces cartography and the display of quantitative geospatial data. The sixth chapter, “*Data Models*” digs into the details of spatial data representation seen in Chapter 1 and an introduction to linking spatial and tabular data. Chapter 7—*Web GIS* has some interesting examples of how web maps actually work but also introduces higher-level applications, including Esri’s ArcGIS Story Maps. Chapter 8—*Open-source GIS* introduces common open-source platforms such as QGIS and GRASS. Chapter 9—*Introduction to Remote Sensing and GIS* covers the basics of remote sensing, from the physics to the platforms, and how remote sensing data is used in GIS. Chapter 10—*Introduction to Health GIS Applications* exposes the reader to public health and emergency management applications and touches on map accessibility. “*GIS and Digital Humanities*” is the title of Chapter 11, describing what the authors mean by the term with some applications of GIS technology to historical and cultural analysis and provides some examples of special challenges of getting essentially non-spatial data into a form that lends itself to GIS analysis. The final chapter, “*Organizational GIS*” talks about how GIS can be used across an organization. This includes discussions of real-time data, data sharing, and integration that is commonly known as “Enterprise GIS.”

The book emphasizes Esri’s ArcGIS platform in its discussion



## GIS: An Introduction to Mapping Technologies, First Edition

Patrick McHaffie, Sungsoon Hwang, and Cassie Follett.

CRC Press, Taylor & Francis Group, Boca Raton, FL. 2019. xiii and 347 pp., diagrams, maps, photos, images, index. Hardcover. \$39.98-105.00. ISBN-13 978-1-4987-4023-4 (hardbound, also available as e-text).

**Reviewed by** Curtis V. Price, Lecturer, Department of Geology and Geological Engineering, South Dakota School of Mines & Technology, Rapid City, South Dakota.

and examples, but it also provides information about other desktop, server, and cloud-based platforms. Surprisingly, although this book was published in 2019, the exercises and tutorials use Esri’s ArcMap 10.5 desktop software and the book has no mention of Esri’s newest desktop application,

Photogrammetric Engineering & Remote Sensing  
Vol. 87, No. 4, April 2021, pp. 229–230.  
0099-1112/21/229–230

© 2021 American Society for Photogrammetry  
and Remote Sensing

doi: 10.14358/PERS.87.4.229

ArcGIS Pro, released in 2015 and quickly gaining adoption. However, an instructor using this book could easily adapt the short and illustrative exercises to whatever software they are using and the content earlier in the chapter would successfully provide the context the student would need to connect the concepts to the user interface (what I call “buttonology”), which can change with every software update. The universal concepts of geography and database management that are the underpinnings of all complete GIS systems are well covered, which makes the book much more “software-agnostic” than it appears at first glance.

The greatest strength of this book (and maybe its biggest weakness) is the ambitious amount of material included. The scope of this small-format (7.5” x 9.5”) book is a little overwhelming; and yet there are a variety of useful treasures that one would not expect in a comprehensive survey of the GIS field. For example, there is a valuable deep dive into a web browser’s HTML source view demonstrating how tiled map services work, a chapter devoted to open-source GIS software, including QGIS and its links to GRASS tools, and another chapter that covers remote sensing in much more

detail than one would expect in a book about GIS. Although though the overall discussion is broad, the exercises in this book commendably “get into the weeds” of what you have to do to import, interpret, and edit data sets found online to make it useful for analysis. This approach may help the reader build a toolbox of technical hands-on methods that are often glossed over in many GIS texts that use pre-formatted included tutorial data in the interest of limiting the scope and simplifying the discussion. Useful details about file formats and data “street smarts” that are often only learned by hard experience are found everywhere in this book. At the same time, there is much here about the philosophy and purpose of geospatial analysis (the “why” of GIS) that I found very inspiring and thought-provoking reading. This book is very hard to classify and there is something here for just about everyone.

Because of its wide scope, I believe this book would be an excellent supplementary text used across many geospatial courses and a great source for the professional to broaden their skillsets and background. I personally plan to spend more time with this book, even after 30 years in the field!



**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.”**

# ASPRS AERIAL DATA CATALOG

**“THE SOURCE FOR FINDING AERIAL COLLECTIONS”**

- 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 for free!
- 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 Contact:**

David Ruiz druiz@quantumspatial.com 510-834-2001	David Day dday@kasurveys.com 215-677-3119
--	---



# & GRIDS DATUMS

BY Clifford J. Mugnier, CP, CMS, FASPRS

## CO-OPERATIVE REPUBLIC OF GUYANA

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 Co-operative Republic of Guyana was originally printed in 2003 but contains updates to their coordinate system since then.

The region known as Guiana or Guayana is on the northeastern coast of South America. It is comprised of the former British Guiana, now Guyana, the former Netherlands Guiana, now Suriname (*PE&RS*, March 2002), and French Guiana. Originally inhabited by the Surinen Indians, the coast was probably sighted by Christopher Columbus in 1498 and by Ojeda and Vespucci in 1499. Vicente Yáñez de Pinzón in 1500 was the first to sail close along the shore. He entered some of the rivers, and the Oyapock River was at first called the Yáñez Pinzón River by the Dutch, and later called the Vincent Pinzón River by the French. Initially, both Spanish and Portuguese mariners avoided the coast between the Oyapock and Orinoco rivers. In 1597-1598, a Dutch expedition examined the river mouths from the Amazon to the Orinoco. This started a series of colonies settled initially by the Dutch, then by the French and the English that all attempted to simultaneously fight the climate, the jungle, and the Indians. The country was eventually ceded to Great Britain by the Dutch in 1814, and it became the British Guiana Crown Colony in 1831. Guyana achieved independence in 1966, it became a republic in 1970, and it adopted a new constitution in 1980.

Slightly smaller than Idaho, Guyana borders Brazil (1,119 km), Suriname (600 km), and Venezuela (743 km). Its coastline on the Atlantic Ocean is 459 km long; it claims a territorial sea of 12 nautical miles, and it claims mineral rights to 200 nautical miles or to the outer edge of the continental margin. Guyana terrain is mostly rolling highlands with a low, swampy coastal plain and a savanna in the south. The



lowest point is the Atlantic Ocean and the highest point is Mount Roraima (2,835 m) in the Pakaraima Range along the Venezuela-Brazil border.

According to Russell Fox of the Ordnance Survey's International Library, "In the first half of the 20<sup>th</sup> century the Lands and Mines Department had observed some high-order astro fixes and traverses in the coastal zone and along the main river valleys leading to the mining areas in the interior of the northern part of British Guyana. There were also a few railway traverses by other entities, such as the Bauxite Company. The only triangulation appears to have been the British Guyana-Brazil boundary survey of the 1930s. The absence of a national triangulation network resulted from logistical, topographical and economic limitations. In the early 1950s the Directorate of Colonial Surveys (DOS) observed some astro fixes, assessed the existing Lands and Mines Department and Bauxite Company work and agreed

Photogrammetric Engineering & Remote Sensing  
Vol. 87, No. 4, April 2021, pp. 231–232.  
0099-1112/21/231–232

© 2021 American Society for Photogrammetry  
and Remote Sensing

doi: 10.14358/PERS.87.4.231

with them on an adjustment strategy to produce a unified set of coordinates from the disparate (albeit high quality) traverses and astro fixes. The method was to accept certain fixes and traverses and to adjust weaker ones to fit; there was not a single datum station. The system adopted was the British Guyana Grid on the Transverse Mercator projection referenced to the International ellipsoid where  $a = 6,378,388$  and  $1/f = 297$ . Latitude of Origin,  $\phi_0 =$  equator, Central Meridian,  $\lambda_0 = 59^\circ$  West of Greenwich, False Easting = 900,000 feet, False Northing = nil, and the Central Scale Factor,  $m_0 = 0.99975$ . The U.S. Inter American Geodetic Survey observed a Hiran trilateration between Venezuela and Brazil in the late 1960s. DOS decided to use that as the basis of a new national network for Guyana and, in 1971, computed an adjustment based on the 1970 Aerodist values of Hiran stations Atkinson and Rose. DOS called the datum Provisional South American 1956. The International ellipsoid was retained but the UTM Grid replaced the British Guyana Grid. That system was used on all subsequent DOS mapping. Note that the 1978 readjustment of the Aerodist trilateration by Matti Jaakkola was not used by DOS. The Royal Engineers and DOS did a lot of Doppler work, DOS to control the aerial photography for its 1:50,000 and large-scale coastal zone map series.”

According to John W. Hager, now retired from NIMA, the

DOS astro fix at Georgetown Lighthouse was published as a mean of determinations observed in 1926 and 1951 where  $\Phi_0 = 6^\circ 49' 31.12''$  N and  $\Lambda_0 = 58^\circ 09' 52.76''$  West of Greenwich. The International ellipsoid was used at the time, as published in the *Guyana Trig List, 3<sup>rd</sup> edition*, 1967. Hager went on to say, “Five stations were established in Guyana as part of the Hiran net from the Caribbean to Brazil. Only station Eagle was held fixed in the Terra Surveys Limited of Canada network established about 1968. The position of Eagle is on the Provisional South American Datum of 1956 where, for Aerodist Station Eagle,  $\phi_0 = 5^\circ 13' 23.6660''$  N and  $\lambda = 59^\circ 06' 10.0549''$  W.”

The origin of the Provisional South American Datum of 1956 (PSAD56) is at La Canoa, Anzoátegui Province, Venezuela where  $\Phi_0 = 08^\circ 34' 17.170''$  N,  $\Lambda_0 = 63^\circ 51' 34.880''$  W, and the defining azimuth to station Pozo Hondo  $\alpha_0 = 40^\circ 22' 45.96''$  (PE&RS, December 2000). La Canoa is about 657 km from Georgetown Lighthouse, and the Zanderij Datum origin in Suriname (PE&RS, March 2002) is about 361 km from Georgetown Lighthouse.

Of all of the borders of Guyana, only the Brazil-Guyana boundary appears to be currently stable and uncontested. Reading the history of the region is a veritable jumble of squabbles among the European powers over centuries that were reflected in the whole of the Guianas by the British, Dutch, French, Venezuelan, and Portuguese. Of late, there has even been some gunboat diplomacy between Guyana and Suriname over the mineral (oil and gas) resources of the offshore continental shelf. Some of the diplomatic position papers are available over the internet, and make for some fascinating reading about the history of the border negotiations.

The latest transformation parameters available from NIMA regarding the shift from the PSAD56 to the WGS84 Datum in Guyana are  $\Delta X = -298 \text{ m} \pm 6 \text{ m}$ ,  $\Delta Y = +159 \text{ m} \pm 14 \text{ m}$ ,  $\Delta Z = -369 \text{ m} \pm 6 \text{ m}$ , and this is based on a solution of nine points in Guyana. Early in 1997, the U.S. National Geodetic Survey published NAD83 coordinates of several local marks in Guyana that were observed by NGS personnel. Being consistent in the NGS “policy” that I consider to be ill conceived and wasteful of U.S. taxpayer funds, no local datum coordinates were researched by NGS personnel nor published in the NGS data sheets. Considering the expenses of a geodetic expedition, consistently not having “sufficient funds” to research a local survey office nor occupy original datum origin points is a specious argument. Let’s try harder, NGS!

### Guyana Update

Guyana has established eight GPS CORS sites throughout the republic, and during the first phase of the Guyanan government’s infrastructure development project, a geodetic survey was conducted in 2019. The localized Provisional South American Datum of 1956 (PSAD56) was bypassed and a new “grid system” (sic) was established on the WGS84 ellipsoid.

<https://www.rics.org/north-america/news-insight/latest-news/news-opinion/a-grid-for-guyana/>

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 (C<sup>4</sup>G). This column was previously published in PE&RS.

### Legal Notice Request for Letters of Interest CSO # 2366 – Task Order Photogrammetry Services

The Connecticut Department of Transportation is seeking to retain one firm to provide task-based photogrammetry services to support its aerial imaging and mapping programs. More detailed information regarding this assignment can be found at: <https://portal.ct.gov/DOT/Consultant-Selection/Consultant-Selection-Information>.

Connecticut Department of  
Transportation  
An Equal Opportunity/Affirmative  
Action Employer



# GREETINGS FROM THE ASPRS STUDENT ADVISORY COUNCIL (SAC)!

**A**s we continuously network with our ASPRS community, we are taking this month to highlight the resources available to Student members and Chapters associated with Universities within the Potomac Region. The Potomac region includes the states of Delaware, Maryland, North Carolina, and Virginia; the District of Columbia; and select counties in West Virginia and Pennsylvania.

The Potomac Region has previously provided financial support to students through funding for activities and travel expenses. Recently, the Region has formalized the process and expanded the financial supports to Chapters and student members. Students and chapters should be aware that funding amounts that exceed established guidelines may be considered, applications are accepted on a continuous basis, and that all activities should promote growth as a student developing in our field. For more information, please visit: <https://www.asprspotomac.org/awards/>. Additionally, you may direct your questions, comments, or suggestions to the President of the Potomac Region at [president@asprspotomac.com](mailto:president@asprspotomac.com), the Secretary-Treasurer at [secretarytreasurer@asprspotomac.org](mailto:secretarytreasurer@asprspotomac.org), or [www.asprspotomac.org/contact/](http://www.asprspotomac.org/contact/).

<p><b>Student Chapter Activities</b></p> <p>Designed to promote interactive events on camps and virtually</p> <p><b>Example Event</b> West Virginia University hosted a technology fair with drones on campus</p> <p><b>Examples Activities</b> Speaker Fees, Refreshments, Transportation for Speakers, Transportation to Events</p> <p>Estimated Funding = 2 events per year @ \$250 per event</p>
<p><b>Inter-Chapter and Inter-University Activities</b></p> <p>Designed to encourage group activities within ASPRS chapters and across universities</p> <p><b>Example Event</b> Organizing a student session at the ASPRS conference or at another related conference</p> <p><b>Example Activities</b> Fostering Relationships, Exchanging Ideas and Presentations, Networking, Research</p> <p>Supports 2 or more chapters or universities and public events are encouraged</p> <p>Estimated Funding = \$500 per event @ 1 per year</p>
<p><b>Student Membership Fees and Student Renewal Fees</b></p> <p>Designed to assist students with membership fees</p> <p>Newly forming chapters are eligible as well as established chapters. Applicants will need to discuss planned activities and/or current activities and career goals. Please submit chapter packages together, instead of individually.</p> <p>Estimated Funding = \$50 for ten students</p>
<p><b>Professional Transition Support</b></p> <p>Designed to support students graduating and entering professional fields</p> <p>Supports first year of membership. Student applicants should document their involvement in their student chapter</p> <p>Estimated Funding = \$150 for one student</p>

ASPRS STUDENT ADVISORY COUNCIL

YOUSSEF KADDOURA  
COUNCIL CHAIR

EVAN VEGA  
DEPUTY COUNCIL CHAIR

JASON MCGLAUGHLIN  
EDUCATION COUNCILOR

GERARDO ROJAS  
DEPUTY EDUCATION COUNCILOR

TERRA MCKEE  
COMMUNICATIONS COUNCILOR

JEFF PU  
DEPUTY COMMUNICATIONS COUNCILOR

MADISON FUNG  
SOCIAL MEDIA COMMITTEE CHAIR

AUSTIN STONE  
NETWORKING COUNCILOR

VICTORIA SCHOLL  
DEPUTY NETWORKING COUNCILOR

LAUREN MCKINNEY-WISE  
CHAPTERS COMMITTEE CHAIR

DAVID LUZADER  
SAC ECPC LIASON

## JOURNAL STAFF

### Editor-In-Chief

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

### Associate Editors

Rongjun Qin, Ph.D., qin.324@osu.edu  
Michael Yang, Ph.D., michael.yang@utwente.nl  
Petra Helmholtz, Ph.D., Petra.Helmholtz@curtin.edu.au  
Bo Wu, Ph.D., bo.wu@polyu.edu.hk  
Clement Mallet, Ph.D., clemallet@gmail.com  
Jose M. Pena, Ph.D., jmpena@ica.csic.es  
Prasad Thenkabail, Ph.D., pthenkabail@usgs.gov  
Ruisheng Wang, Ph.D., ruishengwang@ucalgary.ca  
Desheng Liu, Ph.D., liu.738@osu.edu  
Valérie Gouet-Brunet, Ph.D., valerie.gouet@ign.fr  
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, np@ipf.tuwien.ac.at  
Jan Dirk Wegner, jan.wegner@geod.baug.ethz.ch  
Hongyan Zhang, zhanghongyan@whu.edu.cn  
Dongdong Wang, Ph.D., ddwang@umd.edu  
Zhenfeng Shao, Ph.D., shaozhenfeng@whu.edu.cn  
Ribana Roscher, Ph.D., ribana.roscher@uni-bonn.de  
Sidike Paheding, Ph.D., spahedin@mtu.edu

### Contributing Editors

#### Highlight Editor

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

#### Feature Articles

Michael Joos, CP, GISP, featureeditor@asprs.org

#### Grids & Datums Column

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

#### Book Reviews

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

#### Mapping Matters Column

Qassim Abdullah, Ph.D., Mapping\_Matters@asprs.org

#### Sector Insight

Lucia Lovison-Golob, Ph.D., lucia.lovison@sat-drones.com  
Bob Ryerson, Ph.D., FASPRS, bryerson@kimgeomatics.com

#### GIS Tips & Tricks

Alvan Karlin, Ph.D., CMS-L, GISP akarlin@Dewberry.com

### ASPRS Staff

#### Assistant Director — Publications

Rae Kelley, rkelly@asprs.org

#### Electronic Publications Manager/Graphic Artist

Matthew Austin, maustin@asprs.org

#### Advertising Sales Representative

Bill Spilman, bill@innovativemediasolutions.com

## ASPRS ANNOUNCES ACHIEVEMENT AWARDS

### Steve Coast Selected For The 2021 ASPRS Outstanding Technical Achievement Award

ASPRS has selected Mr. Steve Coast for its prestigious **Outstanding Technical Achievement Award**, recognizing the game-changing and persistent impact of Open Street Map (OSM). His initiative has grown, since its inaugural release in 2004, from a small group in London to a worldwide community of both users and contributors. Because of Steve Coast's vision then and development over nearly two decades, today millions of people use their own resources to easily acquire and to openly share geospatial data acquired with smart devices, low-cost GPS receivers, and airborne or handheld imaging.

The development of OSM was revolutionary. This innovation cultivated the first wide-spread, crowd-sourced earth-observation data acquisition campaign in the mapping community. Steve Coast's work released a novel way to produce useful maps for the general public as well as in professional applications.

In 2006, the OpenStreetMap Foundation was established to encourage the growth, development and distribution of free geospatial data and provide geospatial data for anybody to use and share anywhere. Industry took notice quickly. In December 2006, Yahoo! confirmed that OSM could use its aerial photography as a backdrop for map production. By 2008, OSM data was available for downloading to GPS receivers, and its popularity further increased. Big companies, such as Apple started to use OSM data. OSM impact continues to expand. The OSM community often surveys areas that can be ignored, as fiscally insignificant to large companies, such as jogging trails, footpaths, bicycle routes, boating areas. With the need for anywhere and anytime navigation, OSM recently added path planning functionality, one of the most frequently used mapping-services on smart devices.

Often, special events, named Mapathons, are used to instruct non-professionals, from young students to senior adults, about how to accurately extract data from imagery using readily available applications, including OSM, and how to increase or update geospatial data in key areas, whether in their local area or in places, fraught by the effects of natural events that cause location and content changes. Other examples of the impact of OSM include:

- Humanitarian OpenStreetMap Team (HOT), a nonprofit organization which was created in the immediate aftermath of the 2010 Haiti earthquake.
- The Missing Maps project, founded by HOT, Medecins Sans Frontieres/Doctors Without Borders, and the American and British Red Cross agencies and which aims to map the world's most vulnerable people.
- YouthMappers, founded in 2014 by faculty from Texas Tech University, The George Washington University, and West Virginia University, with support from the US Agency for International Development's

GeoCenter, and now administered by Arizona State University and organized as a network of chapters on university campuses, run by student leadership under the guidance of university professor mentors with 143 campus chapters as of 2018

- GeoChicas, a group of women who volunteer map in OSM and work to close the significant gender gap within the OSM community by promoting mapping campaigns that address women's issues such as mapping gender violence and by creating more training spaces for women and ensuring harassment-free mapping.
- Crowd2Map Tanzania established in 2015 to improve the rural maps of Tanzania to fight female mutilation and improve development of the region.

Visit the OSM website at [www.openstreetmap.org](http://www.openstreetmap.org) to understand more about the activity that Steve Coast initiated and learn why the organization can claim: "OpenStreetMap is a map of the world, created by people like you and free to use under an open license."

Learn more about Steve Coast from his own description <https://wiki.openstreetmap.org/wiki/User:Steve> or from his *The Book of OSM* published in November of 2015.

Steve Coast is a brilliant visionary, an exceptional professional, a successful entrepreneur, a devoted volunteer to the geospatial science and engineering community worldwide, and a staunch advocate of citizen-science contributions.

## Virginia Norwood Named the 2021 ASPRS Life-Time Achievement Award Recipient

Virginia Norwood is known as the "Mother of Landsat." When interviewed by Laura Rocchio for a NASA Landsat Science article, Ms. Norwood was asked if she is comfortable with this title and she replied, "Yes. I like it, and it's apt. I created it. I birthed it; and I fought for it." In recognition of her contributions to satellite communications and optics for over 40 years, including the design, building, promotion and operation of the first multispectral scanner (MSS) on-board Landsat-1, Ms. Norwood was awarded the 2021 ASPRS Life-Time Achievement Award.

As documented in Rocchio (2020), Ms. Norwood graduated from MIT in 1947 at the age of 20 with a degree in mathematical physics. A year later, she and her husband were hired by the U.S. Army Signal Corps in New Jersey where she worked in weather radar. At the age of 22, Ms.

Norwood received the first of her three first U.S. patents, for the invention of a weather balloon radar reflector that allowed storms to be detected and tracked from a distance. Turning to microwave antenna design, she and her husband moved to California to work at Sylvania Electronic Defense Labs. The following year she was hired by the Hughes Aircraft Company where she worked for 26 years until her retirement in 1990. In her capacity as the company's first female member of the technical staff, she received two more patents for her designs of antennae, communication links and optics. When she became the head of the microwave division, one male engineer quit rather than work for a woman and another told her a woman with children had no business working at Hughes. Fortunately for the global remote sensing community, she turned to NASA's need for spaceborne multispectral imagery of the Earth.

After conducting a users' needs study that included interviewing resource managers, collecting field measurements with agronomists and consulting with laboratories at Purdue University and the

## ASPRS WORKSHOP SERIES



### It's not too late to earn Professional Development Hours

Miss an ASPRS workshop or GeoByte? Don't worry! Many ASPRS events are available through our online learning catalog.

<https://asprs.prolearn.io/catalog>

*Image Priscilla Du Preez on Unsplash.*

Environmental Research Institute of Michigan acquiring airborne multispectral imagery, Ms Norwood compiled the specifications for a scanner that would meet the requirements of a broad range of applications. Although interested in scanner technology, NASA engineers were focused on the Return Beam Vidicon (RBV) as the primary sensor for the first Landsat satellite, then named the Earth Resources Technology Satellite (ERTS). The television-like system that had been previously used in TIROS weather satellites and the Surveyor moon missions was similar to the images from traditional cameras and preferred scientists and cartographers of the U.S. Geological Survey. It was finally decided that the multispectral scanner would be added as an experimental payload to ERTS and Norwood and her team proceeded to modify airborne scanning technology to a spaceborne platform with no gravity or air and the need to transmit data extremely long distances. Working with Hughes engineer, Jack Lansing, Ms Norwood refined existing scanner designs to meet spectral and spatial resolutions that exceeded those of weather satellites to produce 4-band multispectral images extending from blue to near-infrared with a pixel size of 79 m. The rotating scan mirror and her decision to transmit the MSS data digitally to the Earth were questioned by NASA, but Ms. Norwood intended the 6-bit image and calibration data from an internal lamp calibrator on-board the satellite would be transmitted digitally to the ground and calibrated to allow spatial and temporal comparisons. As noted in Roccio (2020), “The MSS radiometrically-corrected digital data pioneered quantitative satellite remote sensing, taking the science of Earth observation digital, where it has stayed ever since.”

ERTS-1, later renamed Landsat-1, was launched on July 23, 1972 with both the RBV and MSS sensors. Two days later the first images were received at the NASA Goddard Space Flight Center and all were amazed by the quality of the MSS images. Within two weeks, the RBV experienced an electrical anomaly that ended its image acquisition and the MSS took over as the primary sensor for moderate-resolution Earth observations.

Roccio, L., 2020. Virginia T. Norwood: The Mother of Landsat, Land Science, National Aeronautics and Space Administration. <https://landsat.gsfc.nasa.gov/article/virginia-t-norwood-mother-landsat>.

## WELCOME SHALINI GOGAWALE!

Shalini is ASPRS’s new Association Coordinator and will be replacing Brenna LeMaire. She comes with over 17 years of operational experience in the collegiate athletics industry. She most recently worked on a special project in the Consulting Department of Postlethwaite & Netterville and we are delighted to welcome her to the team!

## THE ASPRS MID-SOUTH REGION PROUDLY ANNOUNCES THE FORMATION OF THE AUBURN UNIVERSITY STUDENT CHAPTER

The new Chapter was recently approved by the Mid-South Region and the Region Officers Council and was presented to the ASPRS Board of Directors

The Chapter’s faculty advisor is Lana L. Narine, Ph.D., Assistant Professor in the School of Forestry and Wildlife Sciences, Auburn University Auburn, Alabama.

## ASPRS ANNOUNCES HENRY JOHN “HANK” THEISS AS THE 2021 RECIPIENT OF THE PHOTOGRAMMETRIC (FAIRCHILD) AWARD

After earning a PhD from Purdue University in 2000 Dr. Theiss became Chief Scientist in Photogrammetry at Centauri. His work significantly advanced the concept of generic sensor modeling of optical line scanners and frame sequences. He developed a new bistatic SAR sensor model with a simulated data testbed and performed associated geopositioning analysis. He supported the Community Sensor Model Working Group (CSMWG) which fosters collaboration among government, industry, and academia to facilitate standardization, interoperability, and advancement of geopositioning capabilities associated with remote sensing systems.

In 2020 he joined the Center for Advanced Spatial Technologies (CAST) at the University of Arkansas as a Research Associate Professor.

The Photogrammetric Fairchild Award is designed to stimulate the development of the art of aerial photogrammetry in the United States. Practicability is the essence of the Award and is the basis for the review of all candidates.

---

### Ad Index

Connecticut Department of Transportation

| 232

# A Digital Terrain Modeling Method in Urban Areas by the ICESat-2 (Generating precise terrain surface profiles from photon-counting technology)

Nahed Osama, Bisheng Yang, Yue Ma, and Mohamed Freeshah

## Abstract

The ICE, Cloud and land Elevation Satellite-2 (ICESat-2) can provide new measurements of the Earth's elevations through photon-counting technology. Most research has focused on extracting the ground and the canopy photons in vegetated areas. Yet the extraction of the ground photons from urban areas, where the vegetation is mixed with artificial constructions, has not been fully investigated. This article proposes a new method to estimate the ground surface elevations in urban areas. The ICESat-2 signal photons were detected by the improved Density-Based Spatial Clustering of Applications with Noise algorithm and the Advanced Topographic Laser Altimeter System algorithm. The Advanced Land Observing Satellite-1 PALSAR-derived digital surface model has been utilized to separate the terrain surface from the ICESat-2 data. A set of ground-truth data was used to evaluate the accuracy of these two methods, and the achieved accuracy was up to 2.7 cm, which makes our method effective and accurate in determining the ground elevation in urban scenes.

## Introduction

Urbanization is one of the most influential human activities across the world today, affecting the quality of urban life and its sustainable development (Ding *et al.* 2016; Sumari *et al.* 2019; Shao *et al.* 2020). With the development of urbanization in the world, the problems caused by urban expansion

Nahed Osama is with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, 129 Luoyu Road, Wuhan 430079, China; and the National Authority for Remote Sensing and Space Sciences, Cairo, Egypt (nahed.osama@whu.edu.cn, nahed.osama2014@yahoo.com).

Bisheng Yang (corresponding author) is with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, 129 Luoyu Road, Wuhan 430079, China.

Yue Ma is with the School of Electronic Information, Wuhan University, Wuhan 430072, China; the School of Science, University of New South Wales, Canberra, BC 2610, Australia; and the Sino Australian Research Centre for Coastal Management, University of New South Wales, Canberra, BC 2610, Australia.

Mohamed Freeshah is with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, 129 Luoyu Road, Wuhan 430079, China; and the Department of Surveying Engineering, Faculty of Engineering at Shoubra, Benha University, 108 Shoubra St., Cairo 11629, Egypt.

are becoming more and more important and bringing new problems to urban development (Sumari *et al.* 2020), such as urban floods (Shao *et al.* 2019), natural vegetation cover decline (Zhang *et al.* 2020), and arable land loss (Shao *et al.* 2020). On the other hand, urban land use is rapidly changing because of urbanization, which, in turn, will cause the terrain to change. The quantity and distribution of urban expansion will affect urban surface runoff and then affect the quality and quantity of groundwater as well (Shao *et al.* 2020; Zhang *et al.* 2020). Therefore, a precise digital elevation product at a global scale is essential for base map production, geographic information systems (GIS), hydrological applications, disaster management, city planning, terrain analysis, infrastructure design, precision farming, and satellite navigation.

In the past, both traditional field surveying and remote sensing techniques have been used to produce elevation models (Erdogan 2009; Tinkham *et al.* 2013). Today, almost all height models are produced by remote sensing, such as UAVs (Pichon *et al.* 2016; Zhang *et al.* 2020), airborne and spaceborne lidars (Carabajal & Harding 2005; Shao *et al.* 2017; Zhang *et al.* 2019), airborne and spaceborne stereo pairs (Singh *et al.* 2010; Pulighe and Fava 2013; Shao *et al.* 2016), and synthetic aperture radar interferometry techniques (Ba *et al.* 2012). The earlier and the current satellite missions, such as the Space Shuttle Radar Topography Mission (SRTM), the Advanced Space borne Thermal Emission and Reflection Radiometer (ASTER), the Advanced Land Observing Satellite-1 (ALOS), Sentinel-1A, Sentinel-1B, TerraSAR-X, and TanDEM-X, have provided global digital elevation information with different spatial resolutions and vertical accuracy.

Lidar is a recent and widely used method to obtain more accurate elevation measurements (approximately up to 15 cm), so integrating lidar data, such as ICESat-1/ICESat-2 data, with other remote sensing data may achieve better elevation modeling results. Previous studies have integrated ICESat-1 mission height measurements with other sources of data, such as the integration between ALOS PALSAR, X SAR, and SRTM, to produce a digital elevation model, achieving a vertical accuracy of 1.14 m (Julzarika 2017). Ocean gravity has been detected by integrating ICESat-1 with ERS-2 (McAdoo *et al.* 2008) and glacier elevation change in Antarctica by integrating ICESat-1 with ERS-2 (Brenner *et al.* 2007). ICESat-1, SRTM, airborne InSAR data, and airborne lidar data with MODIS imagery have been used to detect surface elevation change in Alaska and Canada

Photogrammetric Engineering & Remote Sensing  
Vol. 87, No. 4, April 2021, pp. 237–247.  
0099-1112/21/237–247

© 2021 American Society for Photogrammetry  
and Remote Sensing  
doi: 10.14358/PERS.87.4.237

for a period of 30 years (Muskett *et al.* 2009). Also, surface elevation measurements from the ICESat-1 have been compared to a high-resolution airborne lidar over the Alaska sea ice, and ICESat-1 measurements agreed with the laser altimetry measurement within 2–9 cm in different areas over Alaska (Kurtz *et al.* 2008).

The ICESat-2 carries the first space-based photon-counting lidar to collect height measurements for different surfaces at a global scale and to detect the temporal and spatial change in elevation (Neuenschwander and Magruder 2019). The Advanced Topographic Laser Altimeter System (ATLAS) onboard the ICESat-2 mission can send 10 000 pulses per second from an altitude of 495 km (Neuenschwander and Pitts 2019). The last-generation ICESat-1 was launched to measure ice sheet mass balance, cloud and aerosol heights in 2003 and was operational for 7 years. The current mission was designed to overcome the limitations of the previous mission (Parrish *et al.* 2019). For comparison, the ICESat-1 has only one laser beam with a footprint 70 m in diameter and an along-track interval between adjacent footprints of 170 m, whereas the ICESat-2 has three pairs of beams with footprints of 17 m and an along-track interval of 0.7 m. The multiple beam pairs in ICESat-2 provide better spatial coverage than the ICESat-1 (Schutz *et al.* 2005). The number and distribution of ATLAS-returned photons (including noise and signal photons) is related to the geometry and reflectance of Earth's surface, scattering and attenuation in the atmosphere, and the solar elevation angle (Neuenschwander *et al.* 2019). However, the raw photons captured by the ATLAS is very noisy. For example, in the daytime, the solar background rate reaches many MHz, which means that the number of the background noise photons is much larger than that of signal photons (Herzfeld *et al.* 2014; Markus *et al.* 2017; Ma *et al.* 2018). Noise photons obscure Earth's features (Parrish *et al.* 2019).

Previous researchers have developed many methods for signal photon detection from noisy raw photon-counting data using the Multiple Altimeter Beam Experimental Lidar (MABEL) data sets (Jasinski *et al.* 2016; Nie *et al.* 2018; Ma *et al.* 2019) and MATLAS data sets (produced by adjusting the MABEL data to be similar to ATLAS/ICESat-2 data) (Zhu *et al.* 2018; Forfinski-Sarkozi and Parrish 2019; Neuenschwander and Pitts 2019). For instance, in sea ice and ice sheet areas, a surface-finding algorithm was proposed to detect the surface profiles (Brunt *et al.* 2014; Farrell *et al.* 2015). Other algorithms were developed to detect surface photons in vegetated and forested regions, such as using an adaptive ellipsoid searching filter established from local principal components (Wang *et al.* 2017), using an adaptive density-based model (Zhang and Kerekes 2015), transforming the point cloud into a two-dimensional grid and using active contours for point cloud segmentation (Awadallah *et al.* 2014), and using a

multistep de-noising algorithm based on localized statistical analysis (Nie *et al.* 2018). However, these methods misclassified canopy photons as ground photons in dense vegetation because the laser cannot penetrate deep into the ground surface. Consequently, only a few photons reflected from the terrain surface in areas with dense vegetation. The original density-based notion of clusters algorithm (DBSCAN) was used to identify similar groups of data in a large spatial data set (Ester *et al.* 1996) and was then modified to detect the signal returns from the raw photon-counting lidar data (Zhang and Kerekes 2015).

Regarding the ICESat-2 data processing, the onboard ATLAS algorithm is the model used by NASA's team to classify the photon events. The assumption of the algorithm is that the sunlight background noise recorded by ATLAS follows a Poisson distribution. Thus, this algorithm uses Poisson statistics to find the photons that do not follow this distribution, which are probably signal photon events (Luthcke and Pennington 2019). Most previous studies have dealt with simulated ICESat-2 data collected by airborne micro-pulse lidars, which have high-density photons compared with ATLAS-collected photons. However, little research has utilized the actual ICESat-2 products in their studies to extract the canopy surface and the ground surface (Neuenschwander and Pitts 2019) (Wang *et al.* 2019), land ice surface (Smith *et al.* 2019), and bathymetry information (Parrish *et al.* 2019).

Here, the integration between ICESat-2 lidar data and ALOS PALSAR data has been used to obtain better digital terrain elevation results. The improved DBSCAN model is proposed to detect the signal photons in urban areas from the ICESat-2/ATL03 raw data set. Then the ALOS PALSAR digital surface model (DSM) is utilized to separate the ground photons from the nonground photons of the ICESat-2 signal photons. The objectives of this study are to explore ICESat-2 performance to model the topography in urban areas and to compare NASA's ATLAS algorithm with the improved DBSCAN algorithm in detecting the ICESat-2 signal photons.

## Study Area and Materials

### Study Area

The study area was chosen in Wuhan City, China, through which many ATLAS ground tracks passed. A sampled ground segment captured at 05:39:18 A.M. (UTC time) or 13:39:18 A.M. (local time) on 29 November 2018 was collected and is illustrated in Figure 1. This segment is ground track 1 right (GT1R) and corresponds to one of the three strong beams of ATLAS. This ground track covers the latitude range of 30.4765 N, 30.4965 N and longitude range of 114.5364 E, 114.5497 E. Based on the China's 2018 land use data, the classification

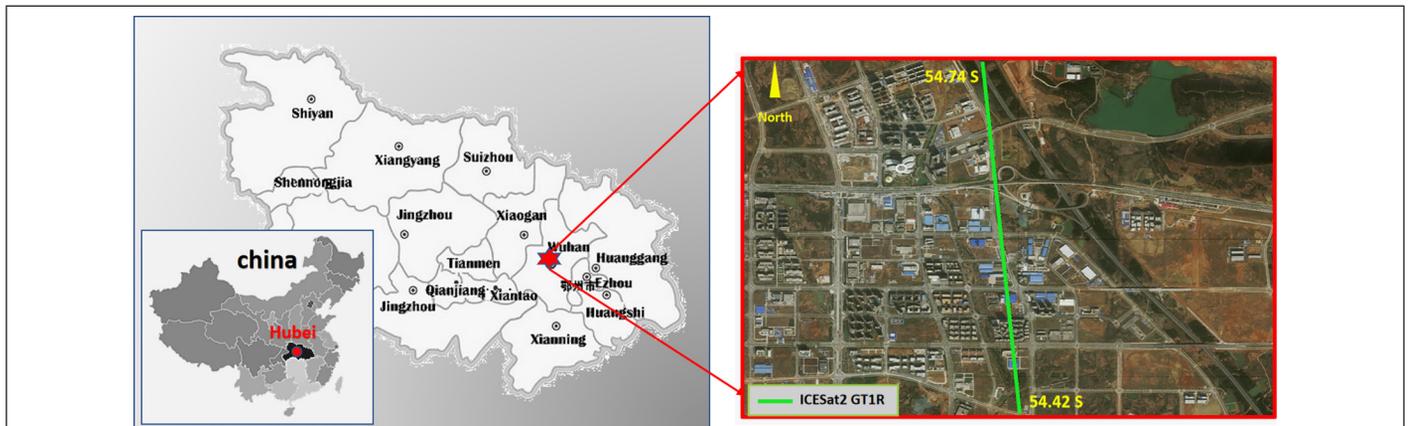


Figure 1. Study area around the Yangtze River in Wuhan, China.

shows that the subset ground track passed through urban and rural settlements and other areas of industrial, mining, transportation, and other kinds of land.

### ICESat-2 Data

ICESat-2 is a recent mission of NASA that was launched at an altitude of 500 km to measure the elevation of Earth's surface (Neuenschwander and Pitts 2019). ATLAS was used to collect height data onboard the ICESat-2 mission. It can take measurements every 0.7 m on Earth's surface. ATLAS has six laser beams that produce six ground segments with a total coverage of approximately 6.5 km in the across-track direction, as shown in Figure 2.

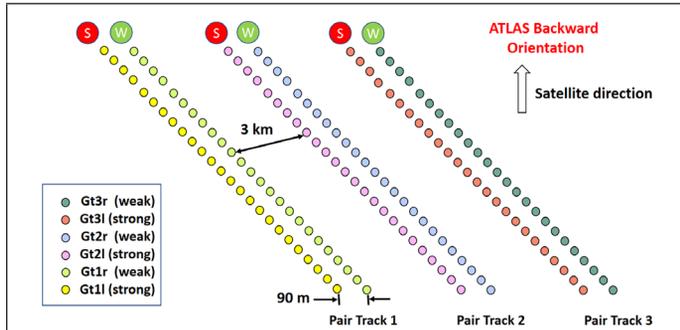


Figure 2. ATLAS ground tracks (strong and weak beams).

ATLAS has many products with different levels of processing. The ATL03 product is a precise global geo-located photon datum that provides accurate values of geodetic latitude, longitude, and ellipsoidal height for each received photon (above the WGS 84 ellipsoid). These photons are corrected from the atmospheric effect and Earth's distortion. The ICESat-2 data is available at <https://icesat-2.gsfc.nasa.gov>.

### ALOS PALSAR Data

PALSAR is one of many instruments onboard the Advanced Land Observing Satellite-1 (ALOS) mission, which was operational from 2006 to 2011. The satellite was launched by the Japanese Aerospace Exploration Agency and can provide radar images for the whole globe. In 2014, the Alaska Satellite Facility has created a project to produce radiometrically terrain corrected (RTC) DSM products by correcting the synthetic aperture radar (SAR) geometry and radiometry (Julzarika 2017). In the radiometric correction, the topographic effect, which changes the backscatter values, was removed. Terrain correction means correcting the geometric distortions that cause some geo-location errors. The correction has been made by integrating many digital elevation models (DEMs) with various accuracies and resolutions. The quality of the RTC product is related to the quality of the DEM used in the RTC process. For this study area and all the regions between latitudes 60 degrees north and 57 degrees south, the SRTM GL1 data at 1-arc-second (30-m) resolution was used for the SAR image correction. The RTC DSM products can be of 12.5 m or 30 m in resolution with a 5-m standard deviation (Santillan and Makinano-Santillan 2016). The DSM used in this research is a high-resolution DSM (12.5 m) that was produced from SAR images captured on 29 January 2009. More information and data download can be obtained from <https://www.asf.alaska.edu>.

### Methods

The ICESat-2 measures photon events at the coordinated universal time (UTC), so the time should be converted first from universal time to local time. UTC is behind the time of Wuhan City by 8 hours, which changes the photon events time from 05:39:18 UTC to 13:39:18 Wuhan. This conversion is important to know if the data were collected in the daytime or the nighttime to set the parameters of the proposed algorithms.

In this study, the ICESat-2 data is daytime data, and here there are many noise photons in the raw point cloud. To obtain the digital terrain model (DTM) from the ICESat-2 data, a new method has been proposed, as shown in Figure 3.

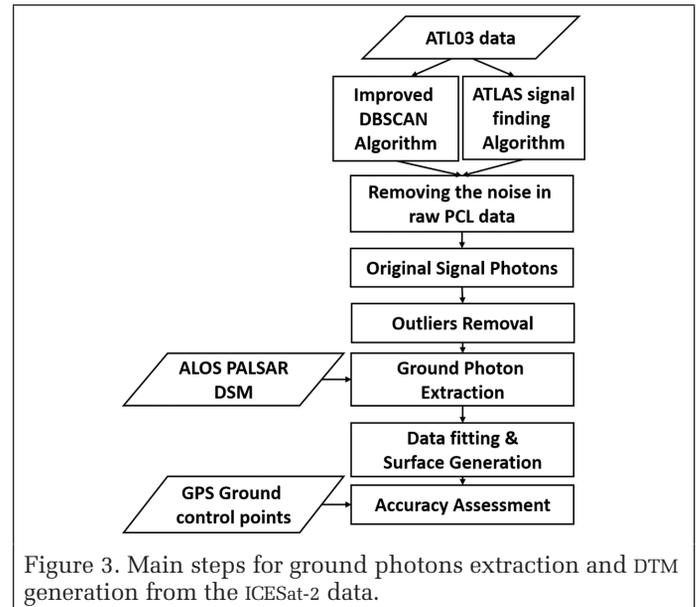


Figure 3. Main steps for ground photons extraction and DTM generation from the ICESat-2 data.

The method depends on the following main steps: (1) Using two different algorithms for de-noising the raw photon-counting data and detecting the signal photon events. In this step, the ATLAS official signal finding algorithm (NASA's algorithm) and the improved DBSCAN algorithm have been used and compared. (2) Removing the remaining noise and smoothing the photons via median filters. (3) Separating the ground photons from the nonground photons using the ALOS PALSAR DSM. (4) Using an averaging window to filter out the ground photons and generate the terrain surface. The next sections explain each step in detail.

### Removing the Noise in Raw PCL Data

The ICESat-2 raw photon point cloud is very noisy (as illustrated in Figure 4) due to the existence of the solar-induced background effect (Markus *et al.* 2017). Therefore, a proper photon classification and extraction algorithm is necessary to detect the signal photons from the raw noisy data.

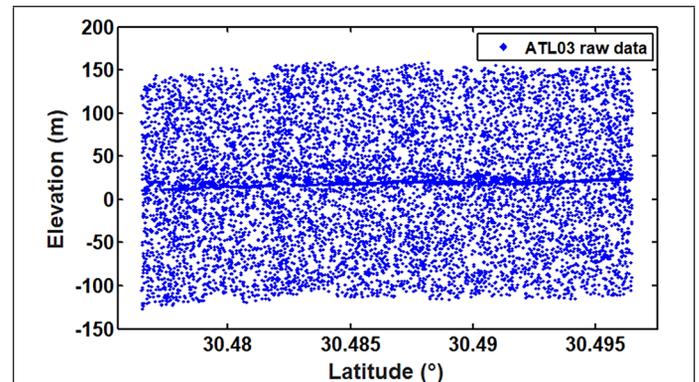


Figure 4. ICESat-2 raw data (ground track 1 right, strong beam).

Detailed information about the two algorithms that have been chosen for the noise removal and signal photon extraction are presented in the following sections.

### Noise Removal by the ATLAS Algorithm

The ATLAS official signal finding algorithm is proposed by NASA's scientists to classify and detect the signal photons collected by the ATLAS device. Given that noise photons satisfy a

Poisson distribution, the algorithm works based on the spatial distribution statistics of raw photons. It classifies the photons into five categories: noise, solar background, low confidence, medium confidence, and high confidence (Scott *et al.* 2019) as illustrated in Figure 5.

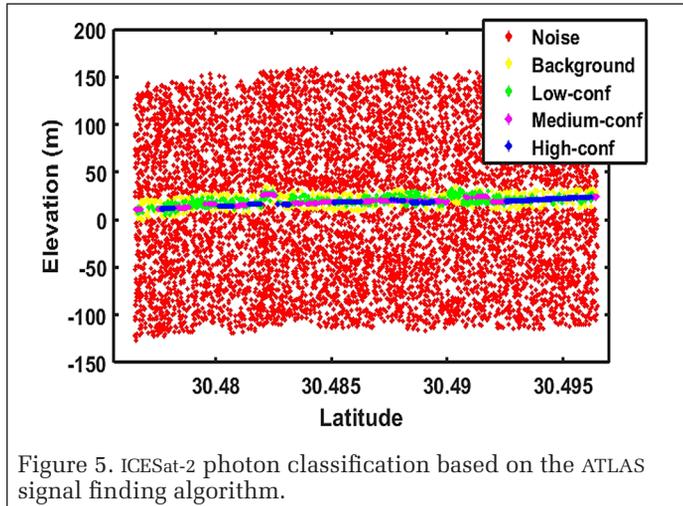


Figure 5. ICESat-2 photon classification based on the ATLAS signal finding algorithm.

The algorithm can distinguish signal photons from other photons by constructing elevation histograms, calculating the signal-to-noise ratio for every bin within this histogram, and determining the confidence parameters based on that ratio. In most cases, the high-confidence photons detected by this algorithm are most likely signal photons.

#### Noise Removal by the DBSCAN Algorithm

Regarding the improved DBSCAN algorithm, this method is a classical clustering method that has been frequently used by previous studies. It can be considered as a supervised classification for the data set segmentation. The principle of this method has been inspired by the ability of the human eyes to observe groups of data by density (Rahmah and Sitanggang 2016). As shown in Figure 6a, we can easily identify the four clusters (C1–C4) of data since the distance between the points is very small. Thus, it has a higher density, and these clusters are separated by bigger distances that make them appear as low-density areas. Therefore, the DBSCAN algorithm groups the data points that are close to each other in one cluster by applying a specific search distance (R) to the neighbor points and detecting the closest neighbor of the core point within the given minimum number of points in each cluster (Min Pts) and the threshold distance R, where the distance between two points  $\leq R$ . Then it groups this point and its neighbors into one cluster and transfers to the next point in the data set to apply the same model for all the data (Martin *et al.* 2009), as shown in Figure 6b.

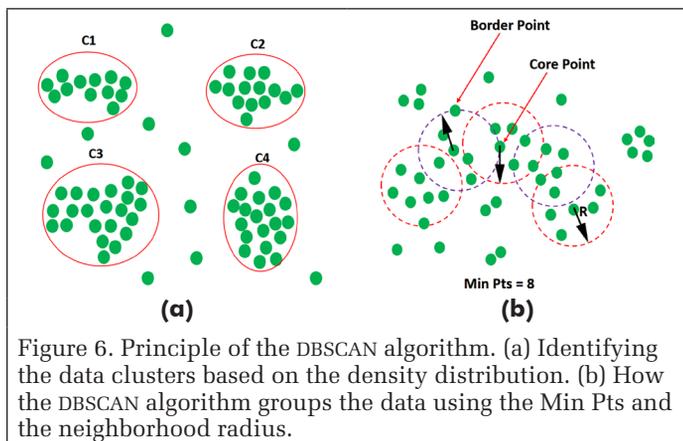


Figure 6. Principle of the DBSCAN algorithm. (a) Identifying the data clusters based on the density distribution. (b) How the DBSCAN algorithm groups the data using the Min Pts and the neighborhood radius.

For estimating the parameters of this model, the minimum number of points was set as three points, as it is not logical to set the Min Pts as one point, and it will be similar to the hierarchical clustering if it is set as two points. Thus, at least three points are required to detect all the possible clusters in the data set. To get the neighborhood distance R, the value can be chosen by constructing the k-distance graph, as proposed in Schubert *et al.* (2017), and the best value will be found in the critical change point in the graph.

Since the improved DBSCAN algorithm detects the photons clusters based on the given neighborhood radius and the minimum number of points in that cluster, it is expected that the algorithm will detect another set of noise clusters with the change of these two parameters. This principle has been used to refilter signal photons after the first detection by applying the algorithm on the data set once again and adjusting the values of R and Min Pts.

#### Outlier Removal

After the application of the improved DBSCAN algorithm, most of the noise photons and the extreme values were removed. Still, some noise photons that exist above and below the ground surface were very hard to be removed without removing some of the important signal photons. Therefore, the median-based Hampel filter and outlier detection filters were used to smooth the data, and great care was taken to preserve the original shape and details of the ground surface. The first filter calculates the median and the standard deviation of the window, which is composed of the sample data, and picks three samples per side surrounded by these data. If the sample exceeds the median by 3-sigmas (standard deviation), it is replaced with the median value (Pearson *et al.* 2016). The second filter is an iterative application of statistical outlier removal as presented by Barnett and Lewis (1994) and the Grubbs test (Grubbs 1969), which tests each sample separately. With a given number of iterations, the tested value is either the maximum value or the minimum one and is the value farthest from the sample mean. For a better application of this filter, the data are preferred to have a normal distribution. That is why this filter should be applied after applying the first filter and removing the probable outliers in the data.

#### Ground Photon Extraction

After filtering and smoothing the data, the remaining photons are expected to be the true signal photons that correspond to the ground surface and aboveground features, such as buildings, trees, and artificial constructions. In previous studies, some methods and algorithms were used to classify signal photons into ground photons, top of canopy photons, and canopy photons (Nie *et al.* 2018; Zhu *et al.* 2018; Narine *et al.* 2019; Neuenschwander & Pitts 2019). However, these methods can be applied in vegetated areas where only the ground surface and trees are involved. In this research, it was difficult to apply the previous methods to effectively separate the aboveground features from the ground surface because the study area contains a mixture of natural and artificial features, especially buildings that mixed with the low-rise trees. Therefore, a method that does not depend on classification algorithms was applied to separate the ground photons from the nonground photons. This method uses an auxiliary source of data that contain the height information of the covered features.

The DTM is a three-dimensional digital representation of bare ground surface, and the DSM is the surface profile that contains the elevation information of covered features above bare ground surface (Sreedhar *et al.* 2015), as illustrated in Figure 7.

It will be expected that if the elevation values of the non-ground features are removed from the ICESat-2 data, the only elevation values that remain will be for the terrain surface (Hu and Tao 2005). There are many methods to extract building information, such as optical remote sensing images (Shao *et*

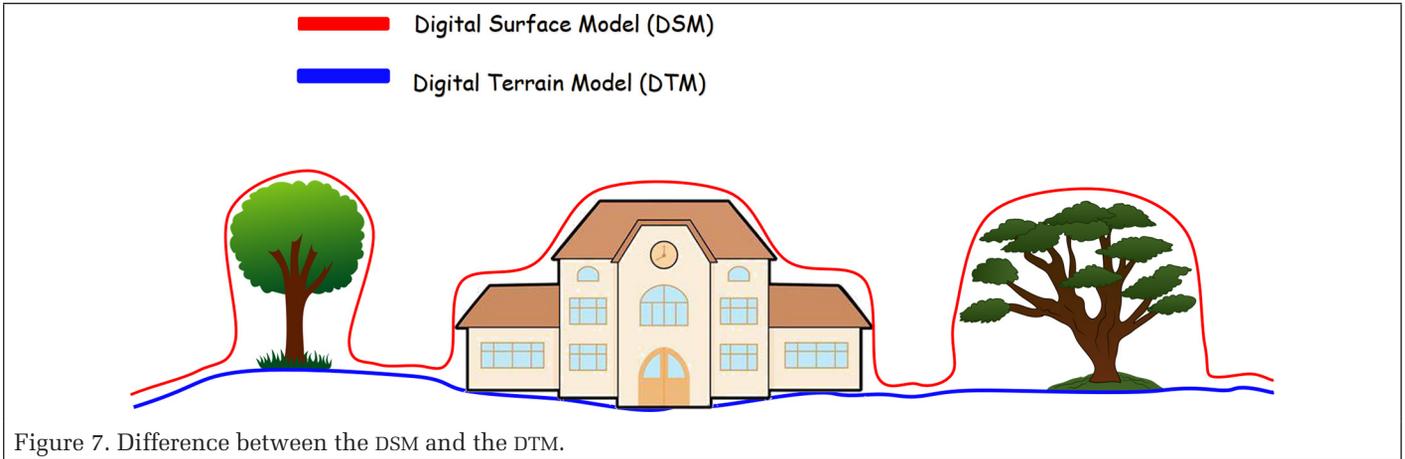


Figure 7. Difference between the DSM and the DTM.

*al.* 2020). However, in order to get accurate results of the bare ground elevations, a surface model that contains both building and trees information, with a high vertical accuracy and a relatively high spatial resolution, should be used. Another dimension in selecting the best DSM for this study is to use open-source data so, it will be easy to apply this method in different study areas around the world. That is why ALOS PALSAR RTC DSM was selected for this purpose. This DSM is considered one of the most accurate high-resolution global open-source DSMs so far. This model is characterized by 5 m of vertical accuracy and 12.5 m of spatial resolution. The DSM is available for download free at <https://search.asf.alaska.edu>.

The following steps were applied to extract height information from the ALOS PALSAR DSM. (1) The area of interest was set, and the RTC DSM for January 2008, which corresponds to the ICESat-2 ground track used in this research, was selected. (2) GIS was utilized to visualize and match the geographic coordinates of the ICESat-2 ground track with ALOS DSM, then the information for the track was extracted using the ArcMap toolbox in the GIS software. The DSM had a shift value and some outliers, that could affect the photon separation process, so it was necessary to remove this shift by taking an intermediate point in the track and subtracting its elevation from the corresponding point elevation in the DSM, then adding this value to all the DSM elevations to eliminate the displacement between the two data sets.

It should be noted that there is a temporal displacement between the ICESat-2 and ALOS DSM because the ICESat-2 data was collected in November 2018 and the ALOS PALSAR images were collected in January 2009. This temporal difference is

also expected to affect the accuracy of ground photon extraction. In order to minimize this effect, it is proposed to divide the study area into relatively small segments (about 1 km each) in the along-track direction, then applying the steps shown in Figure 8 for each segment.

Ground photons will appear continuous and denser than the nonground photons, and by comparing the ALOS DSM with the ICESat-2 elevations, threshold elevation differences were chosen based on the apparent difference between the two elevation sets by the trial-and-error method. Different thresholds for the elevation change should be selected carefully to maintain the ground photons, and the photons above or below these thresholds should be removed. The results at this stage represent only the desired ground photons.

#### Data Fitting and Surface Generation

As a final step to generate the best-fit ground surface, a moving window (at the size of seven) was utilized to estimate the ground surface from the remaining ground photons by taking their average elevations based on this rational transfer function (Oppenheim *et al.* 1999):

$$y(n) = \frac{1}{\text{window size}} (x(x) + x(n-1) + \dots + x(n - (\text{window size} - 1))). \quad (1)$$

This filter minimized the effect of any unremoved surface features, such as small trees and small one-story buildings, that influence the accuracy of the ground surface generation.

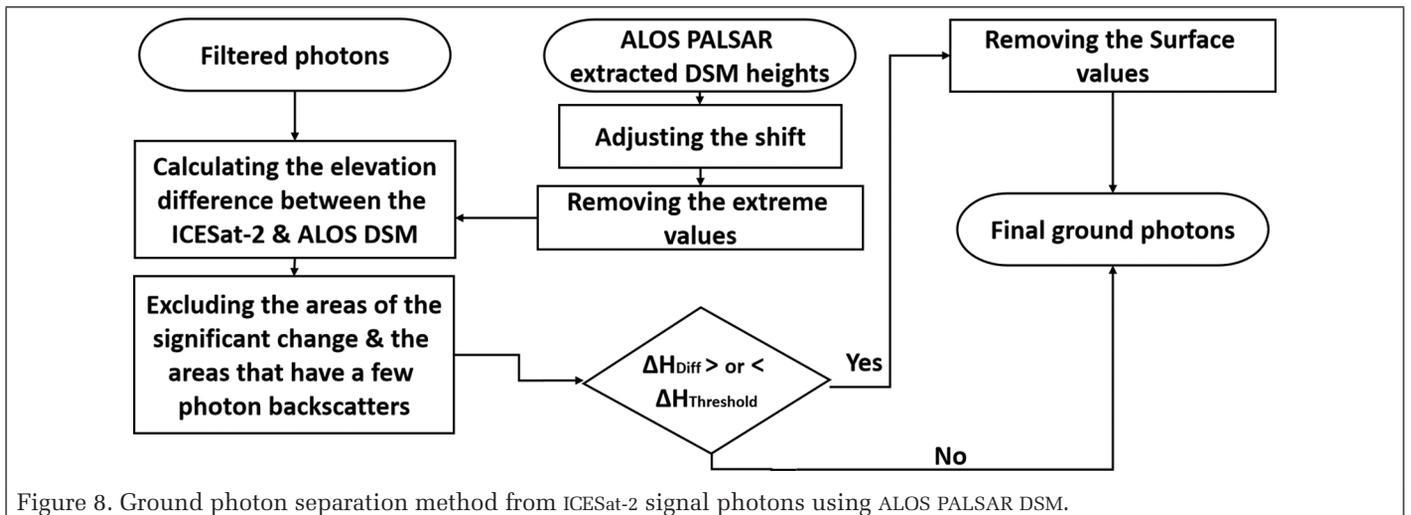


Figure 8. Ground photon separation method from ICESat-2 signal photons using ALOS PALSAR DSM.

## Results and Analysis

### ATLAS Signal Photon Detection Results

The following results were obtained by applying the ATLAS signal finding algorithm. As mentioned, the high-confidence photons are the signal photons of the ICESat-2 data. However, as illustrated in Figure 9a, only a few photons correspond to high confidence, which is not sufficient to recognize Earth's surface. Consequently, the medium-confidence photons were added to ensure that most of the ground photons were included, as illustrated in Figure 9b.

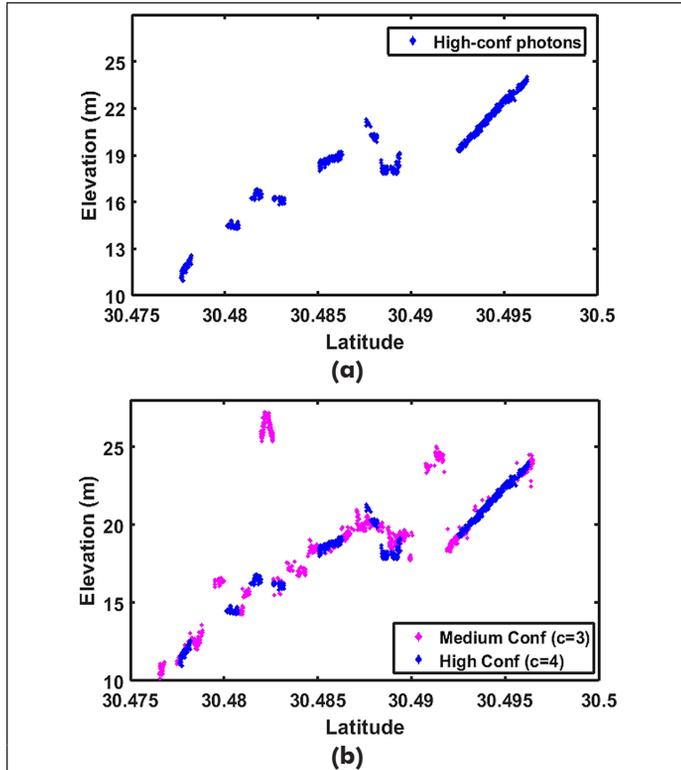


Figure 9. ATLAS signal photons. (a) High-confidence photons. (b) High- and medium-confidence photons.

The advantage of this algorithm is that the classified photons do not have any extreme values and do not need more filters. In Figure 9, it is very clear that the higher the confidence was, the less noise the data contained.

In Figure 9a, the detected photons are probably representing the bare ground surface. However, it also has many gaps when the medium-confidence photons (the purple color in Figure 9b) were added, and some of the above-surface features appeared, such as the symbol “^,” which represents the reflectance of the gable roof of a building, as shown in the green box in Figure 10. Ground photons are missing under this building, as they were obscured by the building.

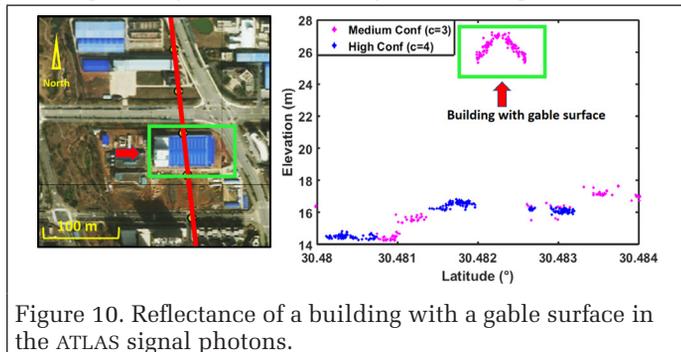


Figure 10. Reflectance of a building with a gable surface in the ATLAS signal photons.

In another case, shown in Figure 11, the ground surface is disconnected due to the existence of some bridges and road intersections in this area that prevent the ICESat-2 laser beam to arrive at the ground surface.

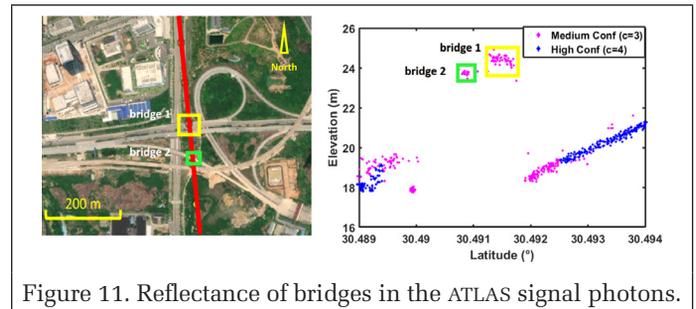


Figure 11. Reflectance of bridges in the ATLAS signal photons.

Slope is an important factor affecting the accuracy (Bolstad and Stowe 1994; Su and Bork 2006; Gillin *et al.* 2015) and reflectance of photon data. The total ground track profile seems to have a gentle slope, and the ground elevation ranges from 10 m to 25 m in an along-track distance of 2500 m.

### DBSCAN Signal Photon Detection Results

For the improved DBSCAN algorithm, the detected signal photons are shown in Figure 12a. Although it looks as if the algorithm efficiently detected the signal photon clusters, the data seem to have some noise clusters too. Thus, the values of two parameters controlling the performance of this algorithm were adjusted; that is, the minimum number of points (Min Pts) was increased, and the neighborhood radius (R) was decreased. Then the DBSCAN algorithm was used again to detect new clusters from the previously filtered data set, as shown in Figure 12b. By the second application, most of the outliers have been removed, and the signal photons were preserved.

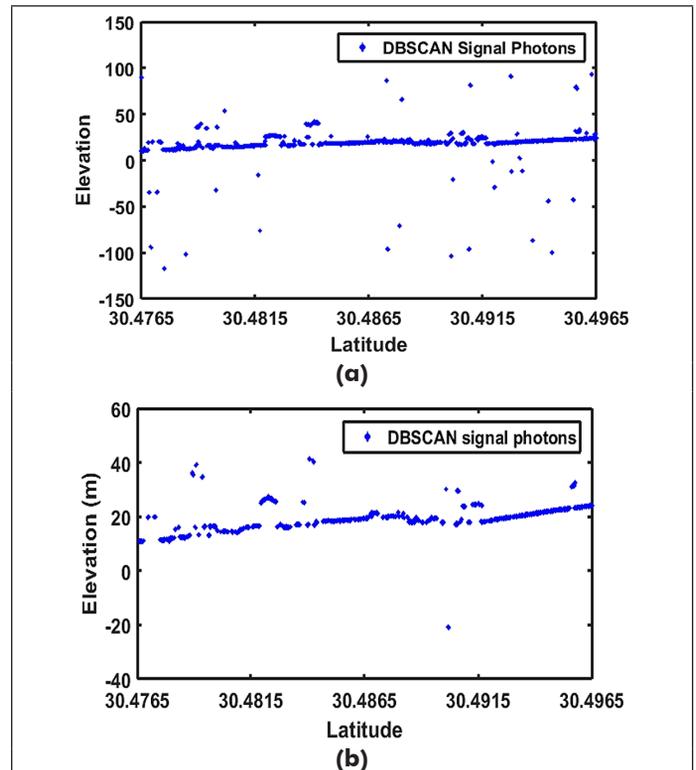


Figure 12. Application results of the improved DBSCAN algorithm on the ICESat-2 raw data. (a) DBSCAN first application. (b) DBSCAN second application.

After applying the DBSCAN for the second time, the data still contain some noise. therefore, a series of iterative filters was used to remove this noise. The filters succeeded in removing the outliers located away from the signal photons. However, the outliers near the ground surface photons were very difficult to recognize and to be removed by any filters, as they were mixed with the ground surface photons.

### Separating Ground Photons by ALOS DSM

After the data filtering, the ALOS PALSAR-derived DSM was used to partially separate the nonground photons from the ground photons by matching the geographic coordinates of the ICESat-2 ground track with the ALOS DSM. Then the surface elevation values for the ground track were extracted using GIS. The ALOS DSM values contained some outliers and was shifted from the ICESat-2 elevations. Thus, the outliers were removed, and the shift distance was adjusted. Then the DSM elevations were compared with the ICESat-2 filtered elevations, as shown in Figure 13.

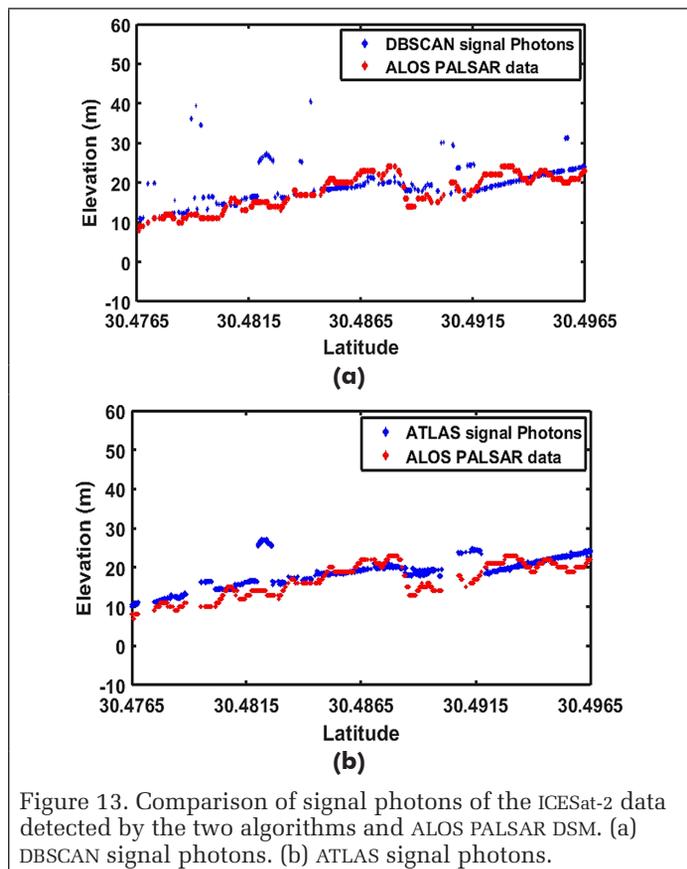


Figure 13. Comparison of signal photons of the ICESat-2 data detected by the two algorithms and ALOS PALSAR DSM. (a) DBSCAN signal photons. (b) ATLAS signal photons.

Since the height accuracy of the ALOS DSM is 5 m, the ALOS DSM did not precisely match the ICESat-2 surface photons and the elevation values distributed above and below that surface. Also, the DSM is greatly mismatched with the ICESat-2 elevations in some areas where the surface shape was changed from the collection date of the ALOS DSM in 2009 to the collection date of the ICESat-2 data in 2018. The ALOS DSM was used to separate the nonground photons from the ground photons by taking the difference in elevations between the two data sets and removing any photons with elevation differences lower than  $-3$  m (lower threshold) or higher than 5 m (upper threshold). These values were chosen after many experiments to prevent any important data loss and to preserve the actual ground photons. After removing the nonground features photons, the DSM values were subtracted from the ICESat-2 values. The results are shown in Figure 14.

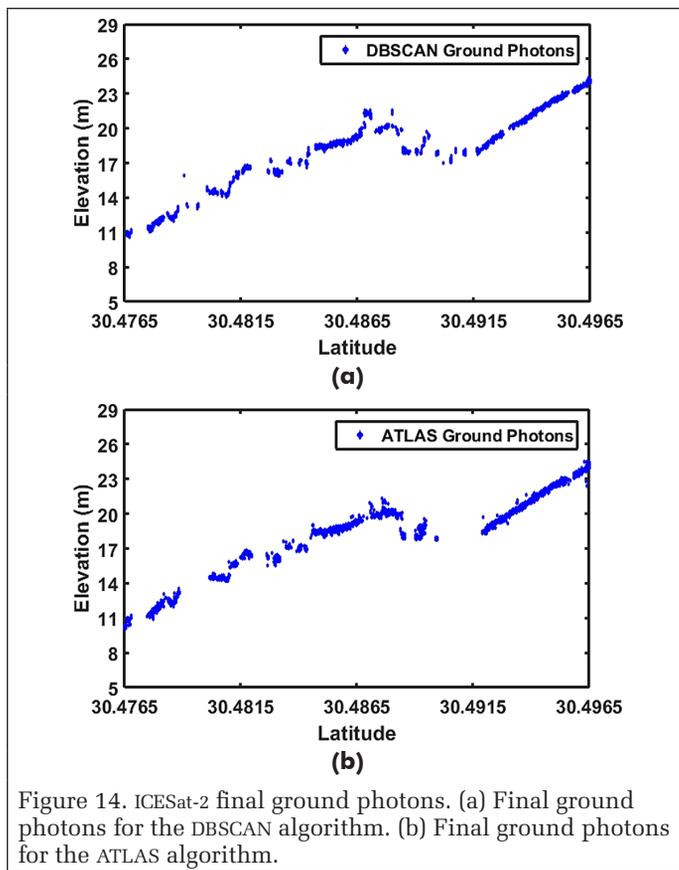


Figure 14. ICESat-2 final ground photons. (a) Final ground photons for the DBSCAN algorithm. (b) Final ground photons for the ATLAS algorithm.

Despite the temporal difference between the ICESat-2 and the ALOS PALSAR data collection time raises some problems in the ground photon separation, the results look promising, and some remaining outliers that were away from the ground surface were removed in this step. The final ground photons look more continuous and connected, indicating that they are the desired ground photons.

### Accuracy Assessment

To evaluate the accuracy of the generated surface, in situ measurements with a higher accuracy were used to evaluate the ICESat-2 results. In this research, a set of eight full ground control points were distributed along the ICESat-2 track, and the precise coordinates of these ground control points based on the WGS-84 ellipsoid were collected on 12 December 2019 using GPS. The points were collected at right-angle intersections with the roads to represent the bare ground elevations. Before performing the statistical test, first, the latitudes and the longitudes of the collected points were matched with the ICESat-2 ground track to ensure that the GPS points lie at the same track and to make sure they are valid to evaluate the results. Second, the points that were located away from the 14-m footprint of the ICESat-2 ground track were eliminated, leaving only five points for the validation process. In this situation, collecting more GPS points was hard due to natural and human constraints, such as when the ICESat-2 ground track passes through private property, government property, or a place that is not suitable to set the GPS instrument or to receive a proper signal. The points were projected in both the ATLAS and the DBSCAN ground photon results after estimating the average ground surface, as shown in Figure 15.

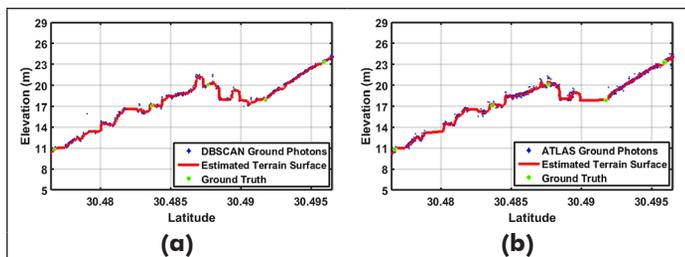


Figure 15. ICESat-2 estimated terrain surface and ground-truth data. (a) Estimated terrain surface for the DBSCAN ground photons. (b) Estimated terrain surface for ATLAS ground photons.

Finally, the vertical accuracy statistic was conducted using the root mean square error (RMSE) test. The results of the test for the two algorithms are listed in Table 1. From Table 1, the mean elevations of the ground levels detected were 17.872 m by the DBSCAN algorithm and 17.914 m by the ATLAS algorithm, respectively. The results from the RMSE test using the GPS ground-truth data indicate that the improved DBSCAN algorithm has a vertical accuracy of 0.027 m, while the ATLAS algorithm has a vertical accuracy of 0.045 m in detecting the terrain surface photons.

Table 1. Accuracy assessment results for the improved DBSCAN and ATLAS algorithms.

Parameter	Mean Elevation of Ground Surface (m)	RMSE (m)
DBSCAN algorithm	17.872	0.027
ATLAS algorithm	17.914	0.045

The elevation differences between the ground-truth data and the improved DBSCAN-detected signal photons range from 2 cm to 6 cm (Figure 16a), whereas the differences were from 0 cm to 15 cm for the ATLAS signal finding algorithm ground photons (Figure 16b).

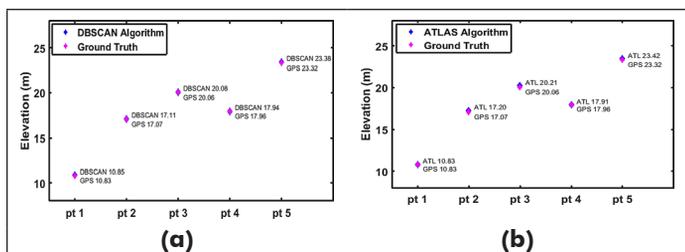


Figure 16. Comparison of the ground-truth elevations data and the corresponding signal photon points detected by the DBSCAN algorithm and the ATLAS signal finding algorithm.

## Discussion

For detecting the signal photons from the ATL03 raw data, the DBSCAN algorithm performed better than the ATLAS signal finding algorithm, and in Table 1, the RMSE difference between the DBSCAN algorithm and the ATLAS algorithm is 1.8 cm. Nevertheless, the results of both algorithms were satisfactory, as both did not exceed centimeter-level accuracy.

### DBSCAN Algorithm

One of the disadvantages of the improved DBSCAN algorithm is that it is not easy to set appropriate values for the two parameters R and Min Pts. Even for the same data set, when the number of samples changes, the parameters should also change (Schubert *et al.* 2017). Also, this algorithm sometimes detects some clusters of the noise photons with the same properties of the signal photons of the Min Pts and R

within each cluster. The results from the algorithm application contain a lot of noise and outliers. Therefore, it had to be removed again by adjusting the number of Min Pts in each cluster and R and using a set of filters to remove the remaining noise and smooth the data. The algorithm is predicted to perform even better in raw data captured at night, when the effect of the solar background is at its minimum and only the signal photons and the instrument noise exist (Zhang and Kerekes 2015).

For the ground surface photons as shown in Figure 14a, the surface detected by the improved DBSCAN algorithm looks more continuous and smoother. Also, the algorithm was able to preserve the surface details even after being applied twice and has better performance in the areas where buildings exist than the ATLAS algorithm. Since the final ground surface was estimated by averaging the ground surface photons, the improved DBSCAN algorithm-estimated ground surface is more accurate than the ATLAS ground surface because when the medium-confidence photons were added to the high-confidence photons in the second method, some noise photons were also included, decreasing the quality of the photons and leading to a worse effect on the surface generation.

### ATLAS Algorithm

On the other hand, regarding the noise filtering, the ATLAS signal finding algorithm seems to have better performance, even with the existence of the background sunlight. The algorithm classification is convenient to use in separating the signal photons from the noise photons without any outliers. In the case of using points of a lower confidence level, surface feature photons appear. However, using the lower confidence points introduces more problems; for example, the noise that is very close to the ground surface and the covered surface features starts to appear gradually. Also, only the high-confidence photons do not provide enough information to represent the ground surface in this research; therefore, we had to add the medium-confidence photons to properly estimate the ground surface. As previously mentioned, the main stumbling block for the ATLAS signal finding algorithm is the slope of the surface (Luthcke *et al.* 2019). That is why the gaps in the signal photons detected by the ATLAS algorithm in this research could be due to the rapid change in the elevations between the ground surface and the top of buildings. Also, the scatter effect of the building's edges makes it hard for this algorithm to detect signal photons on both the ground and the buildings in the intersection areas of the building's edges with the ground surface.

### ALOS DSM

Despite the fact that the ALOS DSM is significantly older than the ICESat-2 data, it was able to filter the nonground photons from the ICESat-2 data and retain the ground surface photons. Nevertheless, the area with great elevation variation experienced some removal of the ground photons as a result of this significant difference in elevations. This could be a disadvantage of this method if a relatively long ground segment was used. Therefore, it is better to choose the appropriate length for dividing the ground segment in the along-track direction based on the apparent elevation change between the two data sets.

The histogram analysis for the improved DBSCAN algorithm and ATLAS signal finding algorithm was performed to evaluate the change in the photon density after conducting each step with both methods. As the signal photons are much closer to each other than the noise photons, a Gaussian curve (Nie *et al.* 2018) can be used to fit the elevation density distribution. Figure 17 shows the fitted elevation density distribution of the photons (using green curves) for the improved DBSCAN algorithm. The area under the green curve is supposed to be the signal photons, and the extension of the red line corresponds to the noise photons and gross errors in the data.

The histogram for the result from the first application of the algorithm is shown in Figure 17a, and the elevations exceed the values of  $-100$  and  $100$ , indicating that the noise still exists above and below the signal photons. After the second application of the algorithm (Figure 17b), the range of the elevation photons decreased from  $-20$  to  $40$ , which ensures that most of the noises were removed by this step. Figure 17c shows the density of the signal photons after the application of the median filter and extracting the ground surface photons from the ALOS PALSAR DSM.

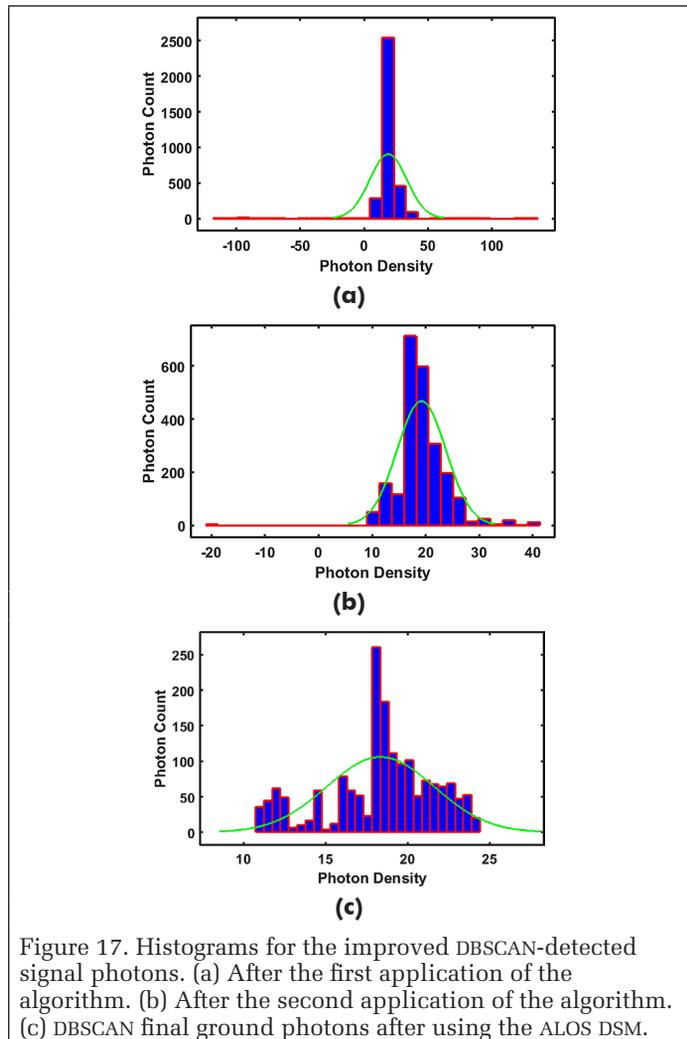


Figure 17. Histograms for the improved DBSCAN-detected signal photons. (a) After the first application of the algorithm. (b) After the second application of the algorithm. (c) DBSCAN final ground photons after using the ALOS DSM.

For ATLAS algorithm histograms, Figure 18 shows the distribution of the high- and medium-confidence photons. It is clear that the photons have a very good distribution and fit the Gaussian curve well in Figure 18a, where almost all the photons lie under the curve. As we have mentioned, this is because the data did not have any extreme values and the noise was at the minimum level because we chose only the high- and medium-confidence photons. This distribution gets even better when we have separate the surface feature photons by ALOS DSM and extracted the ground photons, as the ground photons have a higher density than the nonground photons, as shown in Figure 18b.

When comparing the final histogram of both methods (Figures 17c and 18b), we find that the photon density in the ATLAS algorithm is much higher than that in the improved DBSCAN results, which is also clear in the photon data (Figure 14). The density in the DBSCAN results is highly dependent on the values of the minimum points and the neighborhood radius, which will make the density increase or decrease when increasing or decreasing these two values.

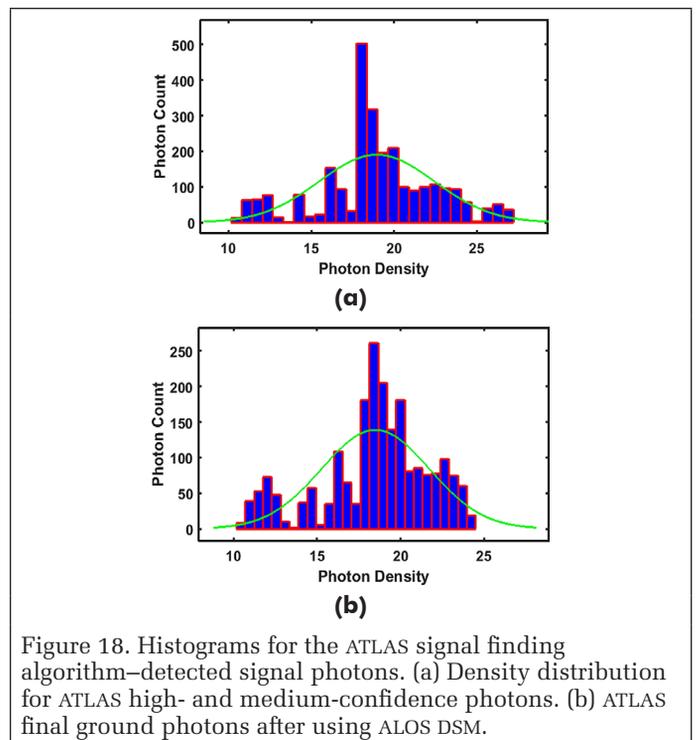


Figure 18. Histograms for the ATLAS signal finding algorithm—detected signal photons. (a) Density distribution for ATLAS high- and medium-confidence photons. (b) ATLAS final ground photons after using ALOS DSM.

One limitation of this method for both the DBSCAN and the ATLAS algorithm is that it will get good results only in relatively small areas (within 5 km) in an along-track distance. For very long tracks, it will take a much longer time to divide the along-track into small segments and choose many thresholds for these segments to subtract the DSM. Despite that, this method is very convenient in the applications that require detailed and high-accuracy information in urban areas.

## Conclusion

The city plays an important role in social development and is the main headquarters for most of the functions of human social, economic, and cultural activities (Li *et al.* 2019). Thus, it is of a great significance to extract or monitor urban surface elevation information by remote sensing.

The ICESat-2 is the latest high-resolution spaceborne photon-counting lidar that is able to map and monitor the change of Earth's elevations precisely through its multiple laser beams with a temporal resolution of 91 days. This study investigated the performance of the ICESat-2 to model the terrain surface using a strong laser beam in urban areas. Two different methods—the improved DBSCAN algorithm and NASA's ATLAS signal finding algorithm—were used to estimate the ground surface elevations. Then the ALOS PALSAR DSM was used to remove the nonground photons. The quality of the two proposed algorithms was evaluated using a collection of GPS ground control points. The evaluation test showed that both methods had good results. Yet the improved DBSCAN algorithm was relatively better than the ATLAS algorithm in estimating the ground surface elevations, with an RMSE value of 2.7 cm. Our method has shown a better result than previous research studies in retrieving the ground photons, making it more convenient to estimate the ground elevations in urban areas.

For future work, the high-resolution full-waveform GEDI (Global Ecosystem Dynamics Investigation) lidar, which will be able to precisely geo-locate the reflectance over different surfaces, is expected to provide more accurate results of the terrain surface by integration with the current ICESat-2 products even in high-density vegetation areas.

## Acknowledgments

This study was jointly supported by the National Science Fund for Distinguished Young Scholars (no. 41725005); the National Key Research and Development Program of China (2018YFB0505401); the Key Research and Development Program of Yunnan province in China (2018IB023); the Research Project from the Ministry of Natural Resources of China under grant 4201-240100123, the National Natural Science Foundation of China under grants 41771452, 41771454, 41890820, and 41901340; and the Natural Science Fund of Hubei Province in China under grant 2018CFA007. The authors would also like to thank Xiaorui Li, who provided expertise that greatly assisted the research.

## References

- Awadallah, M., L. Abbott, and S. Ghannam. 2014. Segmentation of sparse noisy point clouds using active contour models. Pages 6061–6065 in *Proceedings of the 2014 IEEE International Conference on Image (ICIP)*, held in Paris, France, 27–30 October, IEEE.
- Ba, D. N., and G. T. T. Huong. 2012. Comparison of elevation derived from InSAR data with DEM from topography map in Son Dong, Bac Giang, Viet Nam. *Science and Technology* 30 (December):25–29.
- Barnett, V., and T. Lewis. 1994. *Outliers in Statistical Data*, 3d ed. New York: John Wiley & Sons.
- Bolstad, P. V., and T. Stowe. 1994. An evaluation of DEM accuracy elevation, slope and aspect. *Photogrammetric Engineering and Remote Sensing* 60 (11):1327–1332.
- Brenner, A. C., J. P. DiMarzio, and H. J. Zwally. 2007. Precision and accuracy of satellite radar and laser altimeter data over the continental ice sheets. *IEEE Transactions on Geoscience and Remote Sensing* 45 (2):321–331.
- Brunt, K. M., T. A. Neumann, K. M. Walsh, and T. Markus. 2014. Determination of local slope on the Greenland ice sheet using a multibeam photon-counting lidar in preparation for the ICESat-2 mission. *IEEE Geoscience and Remote Sensing Letters* 11 (5):935–939.
- Carabajal, C. C., and D. J. Harding. 2005. ICESat validation of SRTM C-band digital elevation models. *Geophysical Research Letters* 32 (22):1–5.
- Ding, L., Z. Shao, H. Zhang, C. Xu, and W. Dewen. 2016. A comprehensive evaluation of urban sustainable development in China based on the TOPSIS-entropy method. *Sustainability* 8 (8):746.
- Erdogan, S. 2009. A comparison of interpolation methods for producing digital elevation models at the field scale. *Earth Surface Processes and Landforms* 34 (March):613–628.
- Ester, M., H. P. Kriegel, J. Sander, and X. Xu. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. Pages 226–231 in *Proceedings of the 2nd ACM International Conference on Knowledge Discovery and Data Mining (KDD)*, the conference was held in August 2–4, 1996, in Portland, Oregon. Published by The AAAI Press, Menlo Park, California.
- Farrell, S. L., K. M. Brunt, J. M. Ruth, J. M. Kuhn, L. N. Connor, and K. M. Walsh. 2015. Sea-ice freeboard retrieval using digital photon-counting laser altimetry. *Annals of Glaciology* 56 (69):167–174.
- Forfinski-Sarkozi, N. A., and C. E. Parrish. 2019. Active-passive spaceborne data fusion for mapping nearshore bathymetry. *Photogrammetric Engineering and Remote Sensing* 85 (4):281–295.
- Gillin, C. P., S. W. Bailey, K. J. McGuire, and P. Prisley. 2015. Evaluation of lidar-derived DEMs through terrain analysis and field comparison. *Photogrammetric Engineering and Remote Sensing* 81 (5):387–396.
- Grubbs, F. E. 1969. Procedures for detecting outlying observations in samples. *Technometrics* 11 (1):1–21.
- Herzfeld, U. C., B. W. McDonald, B. F. Wallin, T. A. Neumann, T. Markus, A. Brenner, and C. Field. 2014. Algorithm for detection of ground and canopy cover in micropulse photon-counting lidar altimeter data in preparation for the ICESat-2 mission. *IEEE Transactions on Geoscience and Remote Sensing* 52 (4):2109–2125.
- Hu, Y., and C. V. Tao. 2005. Hierarchical recovery of digital terrain models from single and multiple return lidar data. *Photogrammetric Engineering and Remote Sensing* 71 (4):425–433.
- Jasinski, M. F., J. D. Stoll, W. B. Cook, M. Ondrusek, E. Stengel, and K. Brunt. 2016. Inland and near-shore water profiles derived from the high-altitude multiple altimeter beam experimental lidar (MABEL). *Journal of Coastal Research* 76:44–55.
- Julzarika, A. 2017. Height model integration using Alos Palsar, X Srtm C, and Icesat/Glas. *International Journal of Remote Sensing and Earth Sciences* 12 (2):107.
- Kurtz, N. T., T. Markus, D. J. Cavalieri, W. Krabill, J. G. Sonntag, and J. Miller. 2008. Comparison of ICESat data with airborne laser altimeter measurements over arctic sea ice. *IEEE Transactions on Geoscience and Remote Sensing* 46 (7):1913–1924.
- Li, D., J. Ma, T. Cheng, J. L. V. Genderen, and Z. Shao. 2019. Challenges and opportunities for the development of MEGACITIES. *International Journal of Digital Earth* 12 (12):1382–1395.
- Luthcke, S. B., T. Pennington, T. Rebold, and T. Tomas. 2019. Algorithm Theoretical Basis Document (ATBD) for ATL03g ICESat-2 Receive Photon Geolocation. ICESat-2, data products. <<https://icesat-2.gsfc.nasa.gov/science/data-products>> Accessed 20 March 2020.
- Ma, Y., S. Li, W. Zhang, Z. Zhang, R. Liu, and X. H. Wang. 2018. Theoretical ranging performance model and range walk error correction for photon-counting lidars with multiple detectors. *Optics Express* 26 (12):15924.
- Ma, Y., W. Zhang, J. Sun, G. Li, X. H. Wang, S. Li, and N. Xu. 2019. Photon-counting lidar: An adaptive signal detection method for different land cover types in coastal areas. *Remote Sensing* 11 (4):193–210. doi:10.3390/rs11040471.
- Markus, T., T. Neumann, A. Martino, W. Abdalati, K. Brunt, B. Csatho, S. Farrell, H. Fricker, A. Gardner, D. Harding, M. Jasinski, R. Kwok, L. Magruder, D. Lubin, S. Luthcke, J. Morison, R. Nelson, A. Neuenschwander, S. Palm, and J. Zwally. 2017. The Ice, Cloud, and land Elevation Satellite-2 (ICESat-2): Science requirements, concept, and implementation. *Remote Sensing of Environment* 190:260–273.
- McAdoo, D. C., S. L. Farrell, S. W. Laxon, H. J. Zwally, D. Yi, and A. L. Ridout. 2008. Arctic Ocean gravity field derived from ICESat and ERS-2 altimetry: Tectonic implications. *Journal of Geophysical Research: Solid Earth* 113 (5):1–12.
- Muskett, R. R., C. S. Lingle, J. M. Sauber, A. S. Post, W. V. Tangborn, B. T. Rabus, and K. A. Echelmeyer. 2009. Airborne and spaceborne DEM- and laser altimetry-derived surface elevation and volume changes of the Bering Glacier system, Alaska, USA, and Yukon, Canada, 1972–2006. *Journal of Glaciology* 55 (190):316–326.
- Narine, L. L., S. Popescu, A. Neuenschwander, T. Zhou, S. Srinivasan, and K. Harbeck. 2019. Estimating aboveground biomass and forest canopy cover with simulated ICESat-2 data. *Remote Sensing of Environment* 224 (January):1–11.
- Neuenschwander, A. L., and L. A. Magruder. 2019. Canopy and terrain height retrievals with ICESat-2: A first look. *Remote Sensing* 11 (14):1–14.
- Neuenschwander, A., and K. Pitts. 2019. The ATL08 land and vegetation product for the ICESat-2 Mission. *Remote Sensing of Environment* 221 (April 2018):247–259.
- Nie, S., C. Wang, X. Xi, S. Luo, G. Li, J. Tian, and H. Wang. 2018. Estimating the vegetation canopy height using micro-pulse photon-counting LiDAR data. *Optics Express* 26 (10):A520.
- Oppenheim, A. V., R. W. Schaffer, and J. R. Buck. 1999. *Discrete-Time Signal Processing*, 2d ed. Upper Saddle River, NJ: Prentice Hall.

- Parrish, C. E., L. A. Magruder, A. L. Neuenschwander, N. Forfinski-Sarkozi, M. Alonzo, and M. Jasinski. 2019. Validation of ICESat-2 ATLAS bathymetry and analysis of ATLAS's bathymetric mapping performance. *Remote Sensing* 11 (14):1634.
- Pearson, R. K., Y. Neuvo, J. Astola, and M. Gabbouj. 2016. Generalized Hampel Filters. *EURASIP Journal on Advances in Signal Processing* 87 (2016).
- Pichon, L., A. Ducanhez, H. Fonta, and B. Tisseyre. 2016. Quality of digital elevation models obtained from unmanned aerial vehicles for precision viticulture. *OENO One* 50 (3):101–111.
- Pulighe, G., and F. Fava. 2013. DEM extraction from archive aerial photos: Accuracy assessment in areas of complex topography. *European Journal of Remote Sensing* 46 (1):363–378.
- Rahmah, N., and I. S. Sitanggang. 2016. Determination of optimal epsilon (Eps) value on DBSCAN algorithm to clustering data on peatland hotspots in Sumatra. In *IOP Conference Series: Earth and Environmental Science, Vol. 31, Workshop and International Seminar on Science of Complex Natural Systems*, held in Bogor, Indonesia 9–10 October 2015., doi:10.1088/1755-1315/31/1/012012.
- Santillan, J. R., and M. Makinano-Santillan. 2016. Vertical accuracy assessment of 30-M resolution ALOS, ASTER, and SRTM global DEMs over Northeastern Mindanao, Philippines. Pages 149–156 in *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XLI-B4,— XXIII ISPRS Congress*, held in Prague, Czech Republic, 12–19 July 2016, <<https://ui.adsabs.harvard.edu/abs/2016ISPAr41B4..149S>>, provided by the SAO/NASA Astrophysics Data System.
- Schubert, E., J. Sander, M. Ester, H. P. Kriegel, and X. Xu. 2017. DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN. *ACM Transactions on Database Systems* 42 (3):19. <https://doi.org/10.1145/3068335>.
- Schutz, B. E., H. J. Zwally, C. A. Shuman, D. Hancock, and J. P. DiMarzio. 2005. Overview of the ICESat mission. *Geophysical Research Letters* 32 (21). doi: 10.1029/2005GL02400.
- Shao, Z., L. Ding, D. Li, O. Altan, M. E. Huq, and C. Li. 2020a. Exploring the relationship between urbanization and ecological environment using remote sensing images and statistical data: A case study in the Yangtze River Delta, China. *Sustainability* 12 (14):5620.
- Shao, Z., H. Fu, D. Li, O. Altan, and T. Cheng. 2019. Remote sensing monitoring of multi-scale watersheds impermeability for urban hydrological evaluation. *Remote Sensing of Environment* 232:111338.
- Shao, Z., M. E. Huq, B. Cai, O. Altan, and Y. Li. 2020ab. Integrated remote sensing and GIS approach using Fuzzy-AHP to delineate and identify groundwater potential zones in semi-arid Shanxi Province, China. *Environmental Modelling and Software* 134:104868.
- Shao, Z., C. Li, D. Li, O. Altan, L. Zhang, and L. Ding. 2020bc. An accurate matching method for projecting vector data into surveillance video to monitor and protect cultivated land. *ISPRS International Journal of Geo-Information* 9 (7):448.
- Shao, Z., N. S. Sumari, A. Portnov, F. Ujoh, W. Musakwa, and P. J. Mandela. 2020cd. Urban sprawl and its impact on sustainable urban development: A combination of remote sensing and social media data. *Geo-Spatial Information Science*. doi: 10.1080/10095020.2020.1787800.
- Shao, Z., P. Tang, Z. Wang, N. Saleem, S. Yam, and C. Sommai. 2020de. BRNet: A fully convolutional neural network for automatic building extraction from high-resolution remote sensing images. *Remote Sensing* 12 (6):1050.
- Shao, Z., W. Wu, and S. Guo. 2020ef. IHS-GTF: A fusion method for optical and synthetic aperture radar data. *Remote Sensing* 12 (17):2796.
- Shao, Z., N. Yang, X. Xiao, L. Zhang, and Z. Peng. 2016. A multi-view dense point cloud generation algorithm based on low-altitude remote sensing images. *Remote Sensing* 8 (5):381.
- Shao, Z., L. Zhang, and L. Wang. 2017. Stacked sparse autoencoder modeling using the synergy of airborne LiDAR and satellite optical and SAR data to map forest above-ground biomass. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10 (12):5569–5582.
- Singh, V. K., P. K. C. Ray, and A. P. T. Jeyaseelan. 2010. Orthorectification and digital elevation model (DEM) generation using Cartosat-1 satellite stereo pair in Himalayan terrain. *Journal of Geographic Information System* 02 (02):85–92.
- Smith B., H. A. Fricker, N. Holschuh, A. Gardner, S. Adusumilli, K. M. Brunt, B. Csatho, K. Harbeck, A. Huth, T. Neumann, Nilsson, and M. R. Siegfried. 2019. Land ice height-retrieval algorithm for NASA's ICESat-2 photon-counting laser altimeter. *Remote Sensing of Environment* 233:111352.
- Sreedhar, M., S. Muralikrishnan, and V. K. Dadhwal. 2015. Automatic conversion of DSM to DTM by classification techniques using multi-date stereo data from Cartosat-1. *Journal of the Indian Society of Remote Sensing* 43 (3):513–520.
- Su, J., and E. Bork. 2006. Influence of vegetation, slope, and lidar sampling angle on DEM accuracy. *Photogrammetric Engineering and Remote Sensing* 72 (11):1265–1274.
- Sumari, N. S., P. B. Cobbinah, F. Ujoh, and G. Xu. 2020. On the absurdity of rapid urbanization: Spatio-temporal analysis of land-use changes in Morogoro, Tanzania. *Cities* 107:102876.
- Sumari, N. S., G. Xu, F. Ujoh, P. I. Korah, O. J. Ebohon, and N. N. Lyimo. 2019. A geospatial approach to sustainable urban planning: Lessons for Morogoro Municipal Council, Tanzania. *Sustainability* 11 (22):6508.
- Tinkham, W. T., C. M. Hoffman, M. J. Falkowski, A. M. S. Smith, H. Marshall, and T. E. Link. 2013. A Methodology to characterize vertical accuracies in lidar-derived products at landscape scales. *Photogrammetric Engineering and Remote Sensing* 79 (8):709–716.
- Wang, C., X. Zhu, S. Nie, X. Xi, D. Li, W. Zheng, and S. Chen. 2019. Ground elevation accuracy verification of ICESat-2 data: A case study in Alaska, USA. *Optics Express* 27 (26):38168–38179.
- Wang, X., C. Glennie, and Z. Pan. 2017. An adaptive ellipsoid searching filter for airborne single-photon lidar. *IEEE Geoscience and Remote Sensing Letters* 14 (8):1258–1262.
- Zhang, H., X. Ning, Z. Shao, and H. Wang. 2019a. Spatiotemporal pattern analysis of china's cities based on high-resolution imagery from 2000 to 2015. *ISPRS International Journal of Geo-Information* 8 (5):241.
- Zhang, J., and J. Kerekes. 2015. An adaptive density-based model for extracting surface returns from photon-counting laser altimeter data. *IEEE Geoscience and Remote Sensing Letters* 12 (4):726–730.
- Zhang, L., Z. Shao, J. Liu, and Q. Cheng. 2019b. Deep learning based retrieval of forest aboveground biomass from combined LiDAR and Landsat 8 Data. *Remote Sensing* 11 (12):1459.
- Zhang, R., Z. Shao, X. Huang, J. Wang, and D. Li. 2020. Object detection in UAV images via global density fused convolutional network. *Remote Sensing* 12 (19):3140.
- Zhang, Y., and Z. Shao. 2020. Assessing of urban vegetation biomass in combination with LiDAR and high-resolution remote sensing images. *International Journal of Remote Sensing and Earth Sciences* 42 (3):964–985.
- Zhu, X., S. Nie, C. Wang, X. Xi, and Z. Hu. 2018. A ground elevation and vegetation height retrieval algorithm using micro-pulse photon-counting lidar data. *Remote Sensing* 10 (12):1962. doi:10.3390/rs10121962.

## IN-PRESS ARTICLES

- Yang Han, Haiyan Yao, Ziyang Li, Haofang Niu, Tianyi Hao, and Yuyu Zhou. Inversion of Solar-Induced Chlorophyll Fluorescence Using Polarization Measurements of Vegetation.
- Neema Nicodemus Lyimo, Fang Luo, Qimin Cheng, and Hao Peng. Quality Assessment of Heterogeneous Training Datasets for Classification of Urban Area with Landsat imagery.
- Shalini Gakhar and Kailash Chandra Tiwari. Comparative assessment of target detection algorithms for urban targets using hyperspectral data.
- Long Chen, Bo Wu, Yao Zhao, and Yuan Li. A Real-Time Photogrammetric System for Acquisition and Monitoring of 3D Human Body Kinematics.
- Letícia Ferrari Castanheiro, Antonio Maria Garcia Tommaselli, Adilson Berveglieri, Mariana Batista Campos, and José Marcato Junior. Modeling hyperhemispherical points and calibrating a dual-fisheye system for close-range applications.
- Jonathan Li, Qitong Yu, Wei Liu, Wesley Nunes Gonçalves, and José Marcato Junior. Spatial Resolution Enhancement for Large-scale Land Cover Mapping via Weakly Supervised Deep Learning.
- Binbin He, Guoming Li, Liang Li, Miao Yang, Qiongyi Huang, and Zihan Guo. Ecological Functions and Human-activity Interferences Evaluation in Ecological Protection Redline for Urban Environment.
- Fang Gao, Yihui Li, Wentao Li, Peng Zhang, Yuan An, Xing Zhong, and Yuwei Zhai. A High-Resolution Satellite DEM Filtering Method Assisted with Building Segmentation.
- Xueyan Li, Yu Hou, Ruifeng Zhai, Junfeng Song, Xuehan Ma Shuzhao Hou, and Shuxu Guo. Three-dimensional reconstruction of single input image based on point cloud.
- Hui Luo and Nan Chen. A Combined Unmixing Framework for Impervious Surface Mapping on Medium-Resolution Images with Visible Shadows.
- Ze Zhong Zheng, Mujie Li, Mingcang Zhu, Yue He, Jun Xia, Xueye Chen, Qingjun Peng, Yong He, Xiang Zhang, and Pengshan Li. The spatio-temporal evolution of urban impervious surface for Chengdu, China.
- Lei Zhang, Hongchao Liu, Xiaosong Li, and Xinyu Qian. Optimizing the segmentation of a high-resolution image by using a local scale parameter.
- Akib Javed, Qimin Cheng, Hao Peng, Orhan Altan, Yan Li, Iffat Ara, Enamul Huq, Yeamin Ali, and Nayyer Saleem. Review of Spectral Indices for Urban Remote Sensing.
- Qinghong Sheng, Rui Ren, Weilan Xu, Hui Xiao, Bo Wang, and Ran Hong. Spinor-based Attitude Determination with Star Sensor Considering Depth.
- Bharath H Aitha and Prakash PS. Digital building height preparation from satellite stereo images.
- Yanfei Zhong, Xiaoman Li, Yu Su, and Richen Ye. Scene Change Detection based on Multi-feature Fusion Latent Dirichlet Allocation Model for High Spatial Resolution Remote Sensing Imagery.
- Wenhui Wan, Jia Wang, Zhaoqin Liu, Man Peng, Yexin Wang, Tianyi Yu, Chuankai Liu, Lichun Li, and Kaichang Di. Enhanced Lunar Topographic mapping using Multiple Stereo Images Taken by Yutu-2 Rover with Changing Illumination Conditions.
- Jiabao Li, M.D, Wei Han, Ruyi Feng, Lizhe Wang, and Fengpeng Li. Unsupervised representation high-resolution remote sensing image scene classification via contrastive learning convolutional neural network.
- Longjie Ye, Wen Xiao, Yehua Sheng, Dong Su, Pengbo Wang, Shan Zhang, Na Zhao, Hui Chen, and Ka Zhang. Gaussian mixture model of ground filtering based on hierarchical curvature constraints for airborne LiDAR point clouds.
- Jonathan B. Thayn, Aaron M. Paque, and Megan C. Maher. Detecting Geo-Positional Bias in Imagery Collected using Small UASs.
- Chengming Ye, Hongfu Li, Ruilong Wei, Lixuan Wang, Tianbo Sui, Wensen Bai, and Saied Pirasteh. Twice Adaptive Intensity Threshold Method for Uneven LiDAR Data to Extract Road Markings.
- Majid Rahimzadegan, Arash Davari, and Ali Sayyadi. Estimating regional soil moisture with synergistic use of AMSR2 and MODIS images.
- Guoqing Zhou, Man Yuan, Xiaozhu Li, Hongjun Sha, Jiasheng Xu, Bo Song, and Feng Wang. Optimal regularization method based on L-curve for solution of the RFM parameters.
- Lisa M. LaForest, Tian Zhou, Seyyed Meghdad Hasheminasab, and Ayman Habib. System Calibration Including Time Delay Estimation for GNSS/INS-Assisted Push-Broom Scanners Onboard UAV Platforms.

# A Novel Class-Specific Object-Based Method for Urban Change Detection Using High-Resolution Remote Sensing Imagery

Ting Bai, Kaimin Sun, Wenzhuo Li, Deren Li, Yepei Chen, and Haigang Sui

## Abstract

*A single-scale object-based change-detection classifier can distinguish only global changes in land cover, not the more granular and local changes in urban areas. To overcome this issue, a novel class-specific object-based change-detection method is proposed. This method includes three steps: class-specific scale selection, class-specific classifier selection, and land cover change detection. The first step combines multi-resolution segmentation and a random forest to select the optimal scale for each change type in land cover. The second step links multi-scale hierarchical sampling with a classifier such as random forest, support vector machine, gradient-boosting decision tree, or Adaboost; the algorithm automatically selects the optimal classifier for each change type in land cover. The final step employs the optimal classifier to detect binary changes and from-to changes for each change type in land cover. To validate the proposed method, we applied it to two high-resolution data sets in urban areas and compared the change-detection results of our proposed method with that of principal component analysis k-means, object-based change vector analysis, and support vector machine. The experimental results show that our proposed method is more accurate than the other methods. The proposed method can address the high levels of complexity found in urban areas, although it requires historical land cover maps as auxiliary data.*

## Introduction

Urban land cover change detection (CD) has become a hot research issue in recent years. Accurate and detailed urban land cover information is the basis for land use/land cover mapping (Huang *et al.* 2014; M. Lu *et al.* 2016), disaster assessment (Liou *et al.* 2010; Tu *et al.* 2017), urban growth monitoring (Huang *et al.* 2017; Zhang *et al.* 2019), and land management (Vinatier and Arnaiz 2018). The availability of high-resolution (HR) images allows us to perform detailed comparison of urban areas at different phases; however, interaction between humans and natural system complicates land cover change detection in urban areas (X. Wang *et al.* 2018).

In order to extract land cover changes accurately, especially in urban areas, object-based change-detection (OBCD) methods have been developed to take full advantage of spectral information, spatial information, and shape information

(Walter 2004), and they have been shown to be more accurate than pixel-based CD methods for HR images. However, the accuracy of OBCD methods is heavily dependent on the initial segmentation scale (Hussain *et al.* 2013; Tewkesbury *et al.* 2015) and the classification method. Selecting a segmentation scale suitable for urban areas is more complex than for rural areas, because urban land coverages are heterogeneous and changes quickly over time and space.

Selection of the segmentation scale is crucial for subsequent feature extraction and change-information analysis (W. Feng *et al.* 2018). Traditionally, researchers have used simple visual judgment (Meinel and Neubert 2004; Zhang *et al.* 2008; Duro *et al.* 2012) or assessment of segmentation quality (Drăguț *et al.* 2010; Johnson and Xie 2011; Chen *et al.* 2018; T. Feng *et al.* 2018) to select a single scale for segmenting all types of land cover/land use. Land cover changes in urban areas can vary in size, pattern, and complexity, making it extremely difficult to define a scale at which all features in a scene are accurately represented (Johnson and Xie 2011). So recently, researchers have developed multi-scale approaches to address this issue. Some methods integrate multi-scale CD results for greater accuracy (W. Feng *et al.* 2018; Zheng *et al.* 2019). Other methods fuse multi-scale spectral and spatial features to improve CD results (Li *et al.* 2017; X. Wang *et al.* 2018; Shao *et al.* 2019). However, these methods still do not select a scale that intrinsically represents each change type in land cover. Actually, class-specific scale parameters are often necessary when complex land cover changes are encountered (Xun and Wang 2015). Therefore, some methods extract class-specific objects by overlaying historical land cover vector maps with multi-date images (Y. Wang *et al.* 2012; Ayele *et al.* 2018). However, this method assumes that the scale of the vector polygons matches the scale of the change of interest (Tewkesbury *et al.* 2015), so it is unable to detect changes within an object, such as when a part of a building object is changed into a road while the rest of the building object remains unchanged. In addition, Grybas *et al.* (2017) have demonstrated that the optimal parameters for a particular scene do not necessarily equate to the optimal parameters for the chosen classification scheme. So choosing class-specific scale parameters which can highlight local changes and effectively increase the accuracy of CD at the same time is a big challenge.

Not only class-specific scale parameters are needed, but also class-specific classifiers, because different change types in land cover have unique attributes. To date, classifiers such as random forest (RF), support vector machine (SVM), gradient-boosting decision tree (GBDT), and Adaboost have been

---

Ting Bai, Kaimin Sun, Wenzhuo Li, Yepei Chen, and Haigang Sui are with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China (sunkm@whu.edu.cn).

Deren Li is with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China; and the Collaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan, China.

---

Photogrammetric Engineering & Remote Sensing  
Vol. 87, No. 4, April 2021, pp. 249–262.  
0099-1112/21/249–262

© 2021 American Society for Photogrammetry  
and Remote Sensing  
doi: 10.14358/PERS.87.4.249

used to detect changes in land cover type such as buildings, roads, vegetation, bare soil, concrete, and water (Jung 2004; Conchedda *et al.* 2008; Rokni *et al.* 2015). These approaches have high classification accuracy, fast computing speed, and stability in multi-temporal remote sensing classification applications, confirming that different classifiers are suitable for different change types in land cover (Gislason *et al.* 2006; Nemmour and Chibani 2006; Rodríguez-Galiano *et al.* 2011; Verikas *et al.* 2011; Woo and Do 2015; Ustuner and Balik Sanli 2019). Nevertheless, these methods require manual selection of a suitable classifier to achieve high-accuracy CD, and current methods cannot exploit the advantages of all of these methods simultaneously.

To fill this gap, we propose a novel class-specific OBCD method for HR remote sensing imagery specifically designed for complex urban environments. This method has three steps: class-specific scale selection, class-specific classifier selection, and land cover CD. The first step consists of two parts. In the first part, a multi-resolution segmentation method and RF are combined to obtain multi-scale change maps of all types of land cover. In the second part, the harmonic mean of the false-alarm and miss-detection rates exceeding the maximum in these maps acts as a constraint rule in selecting the optimal scale for each change type in land cover. The second step links multi-scale hierarchical sampling (MSHS) with a classifier such as RF, SVM, GBDT, or Adaboost; the algorithm automatically selects the optimal classifier for each change type in land cover. The final step employs the optimal classifier to detect binary changes and from-to changes for each change type in land cover.

The details of this novel class-specific OBCD method are described in the next section. After that are presented the application of the proposed method in urban areas, a performance assessment, and comparative analyses using HR images. Then comes a discussion of the work, followed by conclusions drawn.

## Method

In this study, we propose a novel class-specific OBCD method for urban areas. This method obtains an intrinsic scale and automatically selects a class-specific classifier for each change type in land cover. We provide a flowchart of this method in Figure 1.

The class-specific OBCD method includes three parts: class-specific scale selection, class-specific classifier selection, and land cover CD. Multi-date images and a historical land cover map are first inputted into a class-specific scale-selection method to select an intrinsic scale for buildings, roads, bare soil, water, vegetation, and concrete, as shown at the top of Figure 1. These intrinsic scales are used to segment these land cover types to obtain corresponding objects. A class-specific classifier-selection method automatically selects the optimal

classifier for these classes. The land cover CD uses the optimal classifier to detect binary changes and from-to changes in these classes, as illustrated in the figure. The details of these three parts are given in the following subsections.

## Class-Specific Scale Selection

The first step in the proposed class-specific OBCD method from HR imagery is to determine a class-specific scale. Traditionally, a scale of the vector polygons in the historical vector map represents the scale of the change in land cover (Y. Wang *et al.* 2012; Ayele *et al.* 2018). However, this method cannot detect local changes. Using CD accuracy as a constraint rule to select a smaller scale on the basis of the scale from the vector polygons is a feasible solution. We perform top-down segmentation based on the scale of the vector polygons by multi-resolution segmentation (Gislason *et al.* 2006) to obtain multi-scale objects and use an RF classifier to detect changes. The F1 score, which is the minimum of the harmonic mean of false-alarm rates and missed-detection rates (Beitzel 2006), acts as a constraint rule to select the intrinsic scale for each change in land cover. The flowchart of the proposed class-specific scale-selection method is given in Figure 2.

Suppose there are two remote sensing images: image S1 in T1 and image S2 in T2. These two images are overlaid into

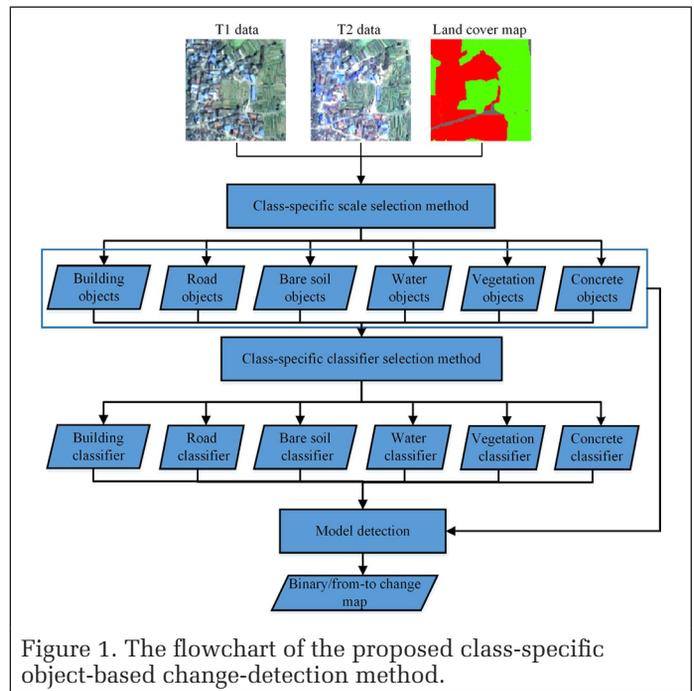


Figure 1. The flowchart of the proposed class-specific object-based change-detection method.

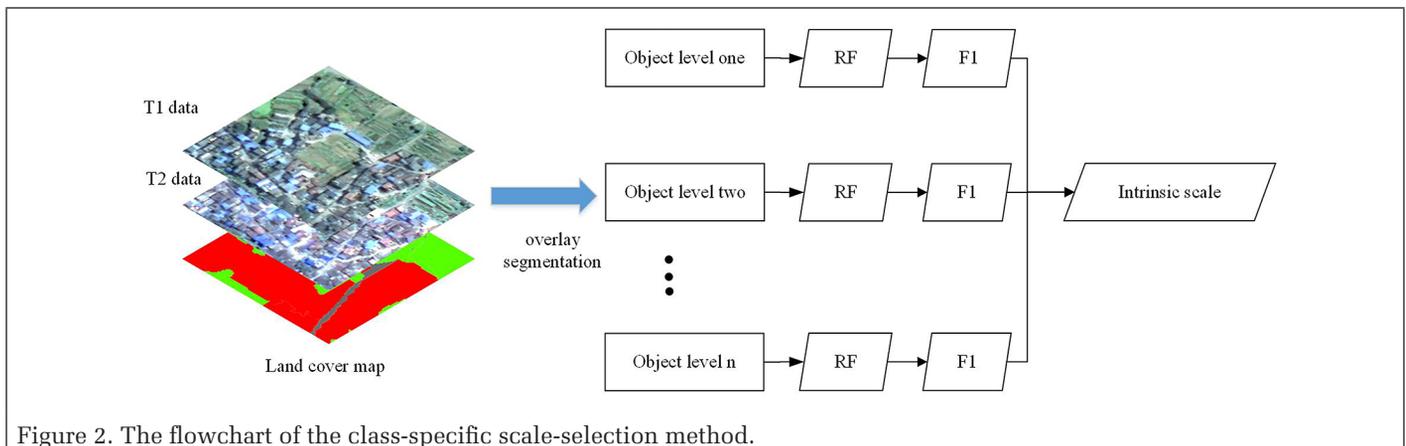


Figure 2. The flowchart of the class-specific scale-selection method.

one image pair  $S$  in Figure 2 by simple band stacking. The multi-resolution segmentation method and a historical land cover map are used to segment this image pair  $S$  to obtain objects. This image pair  $S$  continues to be segmented from top to bottom to obtain subscale object layers  $\{l_1, l_2, \dots, l_n\}$ . After the segmentation process is completed, each object corresponds to a type of change in land cover—including changes in buildings, roads, bare soil, vegetation, water, and concrete—according to the attribute information from the historical land cover map. For each object of image  $S_1$  in  $T_1$  and image  $S_2$  in  $T_2$ , eight types of spectral features (total  $8N$ ), eight types of textural features (total  $8N$ ), and four types of shape features (total  $4N$ ) are selected. Detail information for these spectral, textural, and shape features (Bai *et al.* 2018) of each object in image  $S_1$  and image  $S_2$  is given in Table 1.

As Table 1 shows, all features of two phases are fused to obtain a feature vector ( $36N$ ) for each object. The object feature vectors corresponding to each scale are inputted into RF for CD. The F1 for each change types in land cover at each object level is calculated for verification. The scale corresponding to the largest F1 is the intrinsic scale for each change type in land cover. These intrinsic scales and corresponding subscales will be used as input to the class-specific classifier selection.

### Class-Specific Classifier Selection

HR images reveal various types of change in land cover, each with distinctive imaging attributes. Traditionally, a single-classifier method is used for these change types in land cover (Volpi *et al.* 2013; W. Feng *et al.* 2018). A single-classifier method however, ignores the differences between these change types. To address this issue, we propose a novel class-specific classifier-selection method divided into three parts: basic classifiers, multi-scale hierarchical sampling, and multi-classifier selection.

We first selected four basic classifiers: RF, SVM, GBDT, and Adaboost. We combined MSHS with these four classifiers because these combinations can learn multi-scale features compared with the traditional single-scale sampling method (Bai *et al.* 2018). We conducted multi-classifier selection to obtain a class-specific classifier. In the following subsections, we describe each part in detail.

### Basic Classifiers

Many machine-learning classifiers have been applied to remote sensing areas, including RF, SVM, GBDT, Adaboost, artificial neural networks, deep learning, and so on (Pu *et al.* 2008; Volpi *et al.* 2013; Woo and Do 2015; Gong *et al.* 2016; Bai *et al.* 2018; Georganos *et al.* 2018). From these classifiers, we selected four popular ones—RF, SVM, GBDT, and Adaboost—as the basic classifiers. Deep learning was not considered because it requires a large number of samples. The diagrams of these basic classifiers are given in Figure 3. In the following subsections, we present a brief introduction to each algorithm.

Table 1. Spectral, texture, and shape features for each object.

Feature Type	Features	T1 ( $N$ bands)	T2 ( $N$ bands)	Fusion
Spectral	Mean value for red, green, blue; standard variance for red, green, blue; brightness; and maximum difference	8N	8N	36N
Textural	Mean, variance, homogeneity, contrast, dissimilarity, entropy, angular second moment, and correlation index	8N	8N	
Shape	Area, length/width ratio, shape index, and density	4N	4N	

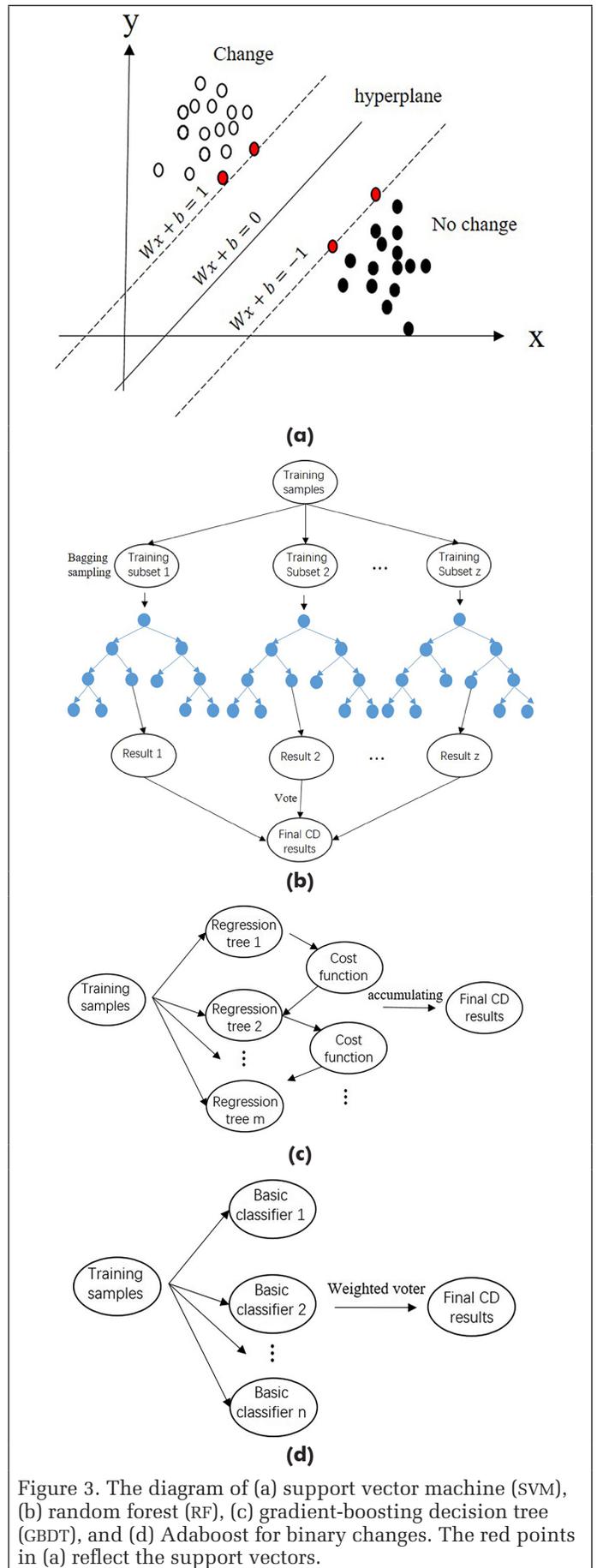


Figure 3. The diagram of (a) support vector machine (SVM), (b) random forest (RF), (c) gradient-boosting decision tree (GBDT), and (d) Adaboost for binary changes. The red points in (a) reflect the support vectors.

One example of SVM, RF, GBDT, and Adaboost can be understood as detecting binary changes between T1 and T2 images. Suppose  $n$  training samples  $T\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$  in this image pair,  $x_i \in X$ ,  $y_i \in \{1, -1\}$ ,  $i = 1, 2, \dots, n$ . The feature space is denoted  $X$ . Each  $x$  in  $X$  has spectral, textural, and shape features. The label of  $x_i$  is denoted  $y_i$ ; if  $y = 1$ , then  $x_i$  has changed—otherwise, no change.

**SVM:** The basic principle of SVM is to find a hyperplane that separates the data set into a discrete predefined number of classes in a fashion consistent with the training samples (Mountrakis *et al.* 2011). As Figure 3a shows, the hyperplane is  $wx + b = 0$ . The distance of the hyperplane from the training sample points  $(x_i, y_i)$  is defined as

$$r_i = y_i \left( \frac{w}{\|w\|} x_i + \frac{b}{\|w\|} \right).$$

Thus the minimum value of the distance of all training sample points from the hyperplane is  $r = \min_{i=1,2,\dots,n} r_i$ , where  $r$  represents the distance from the support vector to the hyperplane. The optimal hyperplane is obtained when  $r$  reaches its maximum value. In SVM, the linear inseparable problem is transformed into a linear classification problem through a kernel function. The radial basis function kernel has demonstrated accuracy in many applications (Anantrasirichai *et al.* 2013; Paneque-Gálvez *et al.* 2013). Therefore, we selected the radial basis function kernel:  $k(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2}$  (Cristianini and Shawe-Taylor 2000).

**Random forest:** RF is composed of a series of classification and regression trees (CARTs), and makes decisions by voting on these decision trees (Breiman 2001). In Figure 3b, the bagging sampling method was used to obtain  $Z$  training subsets from  $T$  and randomly select  $n$  features from spectral, textural, and shape features to split each node, creating  $k$  CARTs:  $f_i(x, \theta_i): x \rightarrow y$ ,  $i = 1, 2, 3, \dots, k$ , where  $x$  is the input vector and  $\theta_i$  is a random vector for all decision trees. This random vector represents a random selection of the training samples or a random selection of features. The final CD results are obtained by voting among these decision trees. The number of decision trees and input features directly decide the final CD accuracy. In general, the number of decision trees is set as 500 (Lawrence *et al.* 2006; Bai *et al.* 2018). The number of features is set to the square root of the number of input features (Gislason *et al.* 2006; Belgiu and Drăguț 2016; Bai *et al.* 2018).

**GBDT:** The GBDT method is a member of the boosting family in ensemble learning. GBDT uses the CART regression tree as a weak classifier, which is also called a multiple additive regression tree or gradient-boosting regression tree (Song *et al.* 2019). In Figure 3c, the cost function is denoted as  $L(y, f(x)) = (y - f(x))^2/2$ . This method obtains the final decision trees by continuously fitting the cost function to the minimum, and produces the final CD results by accumulating the answers of these decision trees.

**Adaboost:** Adaboost was proposed by Freund and Schapire (1997). In Figure 3d, the output weight of the  $k$ th weak classifier is  $W(k) = (w_{k1}, w_{k2}, \dots, w_{kn})$ , where  $w_{i1} = 1/m$ ,  $i = 1, 2, \dots, n$ . So the weighted error rate of the  $k$ th weak classifier  $C_k(x)$  on the training set is  $e_k = p(C_k(x_i) \neq y_i) = \sum_{i=1}^n w_{ki} I(C_k(x_i) \neq y_i)$ , and the weight coefficient of  $C_k(x)$  is

$$\alpha_k = \frac{1}{2} \log \frac{1 - e_k}{e_k}.$$

A weak classifier with a small error rate has a larger weight coefficient. The trials will be terminated when the error rate or trial  $t$  reaches a predetermined value. The final CD result is obtained by the weighted voting of each classifier, according to the function  $f(x) = \text{sign} \left( \sum_{k=1}^K \alpha_k C_k(x_i) \right)$ .

### Multi-Scale Hierarchical Sampling

We combined MSHS with RF, SVM, GBDT, and Adaboost classifiers. The combination of MSHS with the RF classifier can learn multi-scale features from training samples, which yields higher accuracy than the combination of a single-scale sampling method with the RF classifier (Bai *et al.* 2018). Therefore, we use MSHS. Besides the combination of MSHS with RF, we expanded the application scope of the MSHS method, combining it with SVM, GBDT, and Adaboost. An example of a combination of MSHS and SVM is shown in Figure 4.

As Figure 4 shows, for buildings, roads, bare soil, vegetation, water, and concrete objects, the respective training samples at the current intrinsic scale and subscales are inputted into MSHS to obtain different scale-sample combinations. These multi-scale sample combinations are subsequently inputted into multiple SVM classifiers. The cross-validation (CV) error in training samples is calculated. Bai *et al.* (2018) used the minimum out-of-bag error parameter as the constraint rule to select the optimal RF classifier for the combination of MSHS and RF. We used the CV error in training samples instead, for all four classification methods, since SVM, GBDT, and Adaboost have no out-of-bag error parameters, and deployed the minimum CV error in the verification samples to select the optimal SVM classifier. The same extension operation is applied to the RF, GBDT, and Adaboost classifiers, and the optimal classifier in each of the four types is selected. In the following subsection, we discuss the multi-classifier selection process.

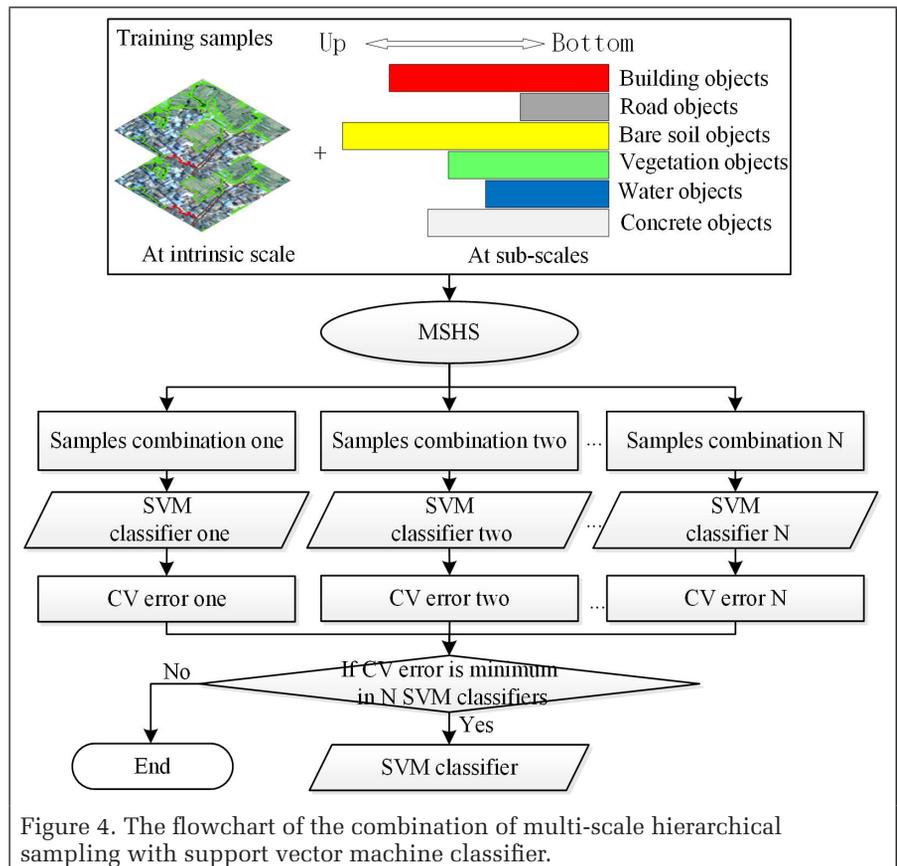


Figure 4. The flowchart of the combination of multi-scale hierarchical sampling with support vector machine classifier.

### Multi-Classifer Selection

In this subsection, we discuss the selection of an optimal classifier for each change type in land cover. An example of the multi-classifier selection method is shown in Figure 5.

Figure 5 illustrates a multi-classifier selection for building objects. The building segmentation scale is used to segment image S1 in T1 and image S2 in T2 to obtain building objects. For building objects of this image pair, eight types of spectral features (mean value for red, green, and blue; standard variance for red, green, and blue; brightness; and maximum difference), eight types of textural features (mean, variance, homogeneity, contrast, dissimilarity, entropy, angular second moment, and correlation index), and four types of shape features (area, length/width ratio, shape index, and density) are selected. All features of two phases for building objects are fused to obtain a feature vector. Building objects and corresponding feature vectors are inputted into the combined MSHS method with RF, SVM, GBDT, and Adaboost classifiers to obtain the optimal SVM, RF, GBDT, and Adaboost classifiers. We select the final classifier for building objects from these optimal classifiers when the CV error of the verification samples exceeds the minimum in these classifiers. The same extension operation is applied to roads, vegetation, bare soil, water, and concrete objects to obtain a final classifier for each type of objects.

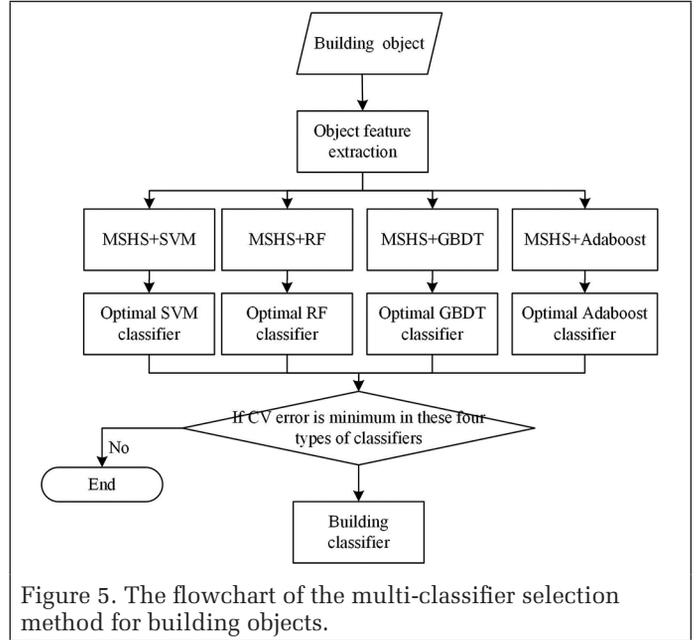


Figure 5. The flowchart of the multi-classifier selection method for building objects.

### Land Cover CD

Land cover CD is the last step of our proposed method. In this step, we employ the final class-specific classifiers to detect binary changes and from-to changes for buildings, roads, bare soil, vegetation, water, and concrete objects.

### Results and Analysis

In this section, we present the experimental data, sampling process, experimental results, and accuracy analyses of the proposed method.

#### Experimental Data

Experimental data were collected based on two multi-temporal remotely sensed images corresponding to the geographical areas of Suizhou and Liuzhou cities, China. Land cover in these two cities has been changing rapidly. Only the red, green, and blue bands of the images were used in our study. A reference change/no-change map and historical land cover maps come from the China Surveying and Mapping Department. All images have a resolution of 2 m. Each image pair was preprocessed using image registration (Townshend *et al.* 1992) and relative radiometric normalization (D. Lu *et al.* 2004).

Data set 1, shown in Figure 6, consists of a land cover map from 2015, a reference change/no-change map, and two multispectral images from the WorldView-2 satellite over an area of Suizhou, acquired in 2015 and 2016. It covers 768×768 pixels and contains six main land cover classes: buildings, roads, bare soil, water, vegetation, and concrete. Reference binary changes for the data set are change or no change.

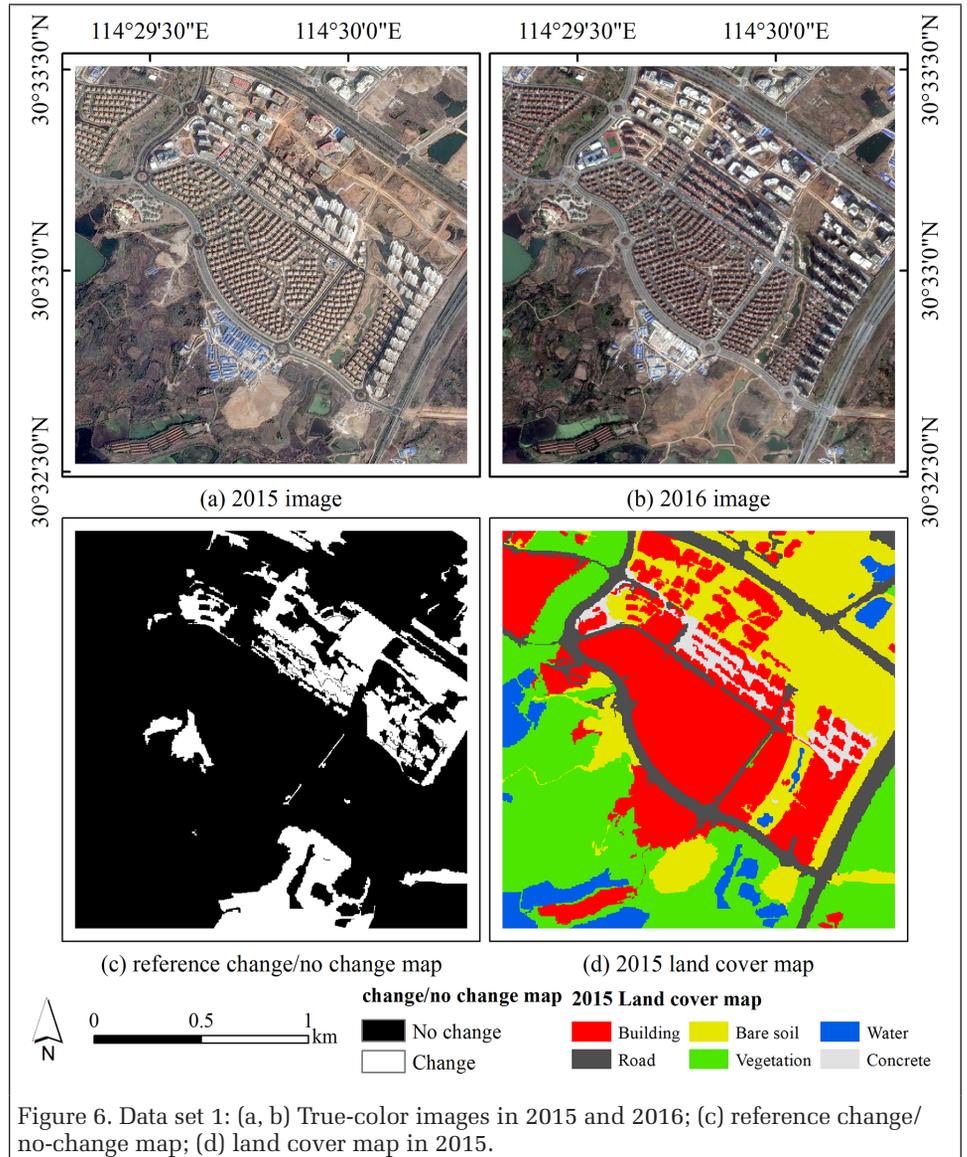


Figure 6. Data set 1: (a, b) True-color images in 2015 and 2016; (c) reference change/no-change map; (d) land cover map in 2015.

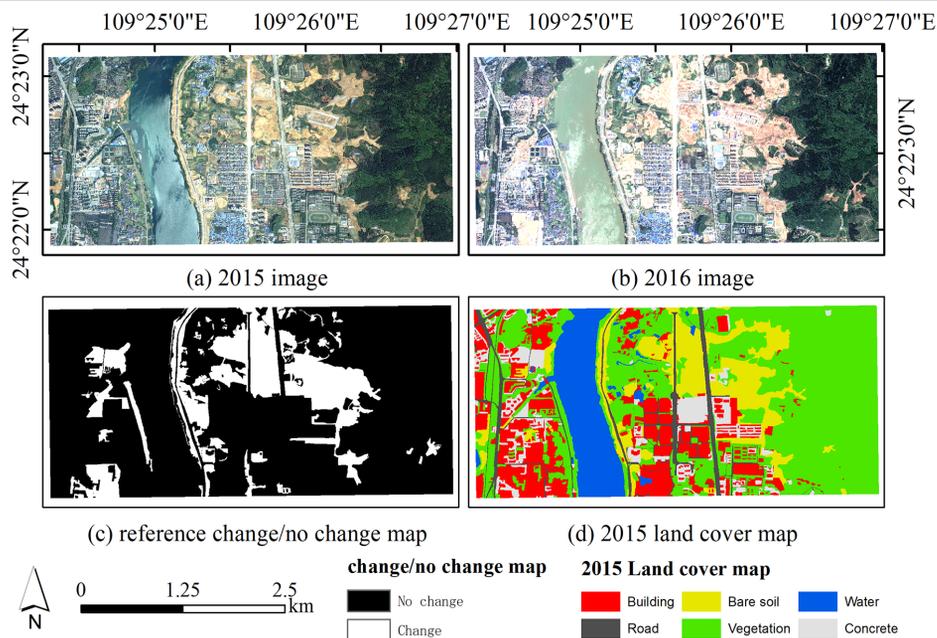


Figure 7. Data set 2: (a, b) True-color images in 2015 and 2016; (c) reference change/no-change map; (d) land cover map in 2015.

Data set 2, shown in Figure 7, consists of a land cover map made in 2015, a reference change/no-change map, and two multispectral images from the Gaofen-2 satellite over an area of Liuzhou acquired in 2015 and 2016. It covers  $2256 \times 1137$  pixels. Like data set 1, data set 2 contains six main land cover classes: buildings, roads, bare soil, water, vegetation, and concrete. Reference binary changes for data set 2 are change or no change. We applied and validated our proposed method using these two data sets as follows.

### Sampling

For sampling, we used 20% training-set ratios to implement simple random polygon sampling of the pixel ground truth. Training samples were obtained manually. They also can be obtained from the pre-change detection map. Pre-change maps can be generated by unsupervised CD methods such as principal component analysis (PCA) *k*-means (Celik 2009), iteratively reweighted multivariate alteration detection (Nielsen 2007), change-vector analysis (Sun and Chen 2010), and so on. Training samples covered all land cover change types evenly. In the experiment testing the optimal segmentation-scale and multi-classifier selection, 75% of the training samples were selected as the training data and 15% were used as the verification data to calculate the CD accuracy. In the CD method experiment, the remaining pixels in images were used as the validation set to calculate the final CD accuracy.

### Experiments

In our experiments, for the RF classifier we set the number of decision trees to 500 and the number of splitting variables to six. For the SVM classifier, we used the radial basis function kernel. The change types in land cover were defined using the land use map from the previous phase, including changes from buildings, roads, bare soil, vegetation, water, and concrete.

In the class-specific scale-selection experiment, we compared the CD results at the 10 segmentation scales. For data set 1, 10 segmentation scales  $\{l_1, l_2, \dots, l_{10}\}$  measured increments of 10 ranging from 10 to 100. For data set 2, 10 segmentation scales  $\{l_1, l_2, \dots, l_{10}\}$  measured increments of 20 ranging from 20 to 200. The training objects at 10 segmentation scales were guaranteed to stay in the same position during sampling. We overlaid random polygon samples with the objects at scale

110 to obtain training objects. The object with the most pixels superimposed on the random polygon sample was selected as the training object. After training objects were determined at scale 110, they were determined at scales 11–19 by the hierarchical relationship between the multi-scale layers. The label of each training object was determined according to the pixel ground truth. Our experiments were performed in Python.

In order to validate this method, we applied our method to the two data sets to detect binary and from-to changes, then compared the binary and from-to CD results using the proposed method with those from the traditional single-scale, single-classifier (SSSC) method and a class-specific-scale, single-classifier (CSSC) method. For the SSSC method, we used a single RF classifier to detect changes at 10 scales from 10 to 100 with increments of 10 for data set 1, and at ten scales from 20 to 200 with increments of 20 for data set 2. For the CSSC method, we used a single RF classifier to detect changes at a class-specific scale. In the following, we will conduct a detailed analysis of these binary and from-to CD results for the two data sets.

#### Binary CD Results

The binary CD experimental results of the two data sets using SSSC methods at 10 single segmentation scales, the CSSC method, and our proposed method are shown in Figures 8 and 9. White regions represent change and black regions represent no change area.

As can be seen from the figures, the binary CD results from small scales to large scales are different for the two data sets. Minor changes can be detected at small scales, but this includes many false alarms. There are many missed detections at large scales. These results demonstrate that the accuracy of CD is heavily dependent on scale. In addition, a visual comparison of results from the proposed method with those from the SSSC and CSSC methods reveals that our proposed method produced is more visually accurate.

#### From-To CD Results

The from-to CD experimental results of the two data sets using SSSC methods at 10 single segmentation scales, the CSSC method, and our proposed method are shown in Figures 10 and 11. Reference from-to change maps for the two data sets were from the China Surveying and Mapping Department.

Reference from-to changes for data set 1 comprise 22 types: no change and 21 types of land cover changes, such as from building to road. From-to changes for data set 2 comprise 27 types: no change and 26 types of land cover changes. Black regions represent no change and colored regions represent from-to change.

From a comparison of our results to the reference ground truth, our proposed method appears more accurate visually than the SSSC methods at 10 scales and the CSSC method. These results suggest that our proposed CD method can still be applied to from-to CD.

### Experimental Analysis

In order to validate our proposed CD method, we selected false-alarm rate (FAR), missed-detection rate (MDR), overall accuracy (OA), and F1 measures (Lunetta *et al.* 2004; Beitzel 2006). These four indexes range from 0 to 1. FAR and MDR get closer to zero with our proposed method, and OA and F1 get closer to 1, illustrating that our proposed method is more accurate.

#### Accuracy Analysis with SSSC and CSSC Methods

FAR, MDR, OA, and F1 values of binary and from-to CD results were calculated for the two data sets for our proposed method versus the SSSC and CSSC methods. Due to spatial limitations, only binary CD accuracy values for the two data sets are shown in the tables.

Values for FAR, MDR, OA, and F1 of binary CD using the proposed method versus the SSSC and CSSC methods for data sets 1 are shown in Table 2. The MDR of our proposed method (16.93%) is significantly lower than for all 10 SSSC scales (42.01%–72.29%). Our proposed method delivers higher OA (93.42%) and F1 (82.27%) than all 10 SSSC scales (OA: 83.13%–89.96%; F1:

Table 2. False-alarm rate (FAR), missed-detection rate (MDR), overall accuracy (OA), and F1 values of binary change-detection results using our proposed method, single-scale, single-classifier (SSSC) methods, and the class-specific-scale, single-classifier (CSSC) method for data set 1.

Method	FAR (%)	MDR (%)	OA (%)	F1 (%)
SSSC				
10	14.54	50.78	89.14	62.47
20	18.09	42.01	89.93	67.90
30	14.77	45.19	89.96	66.72
40	23.38	57.55	87.05	54.63
50	38.42	50.37	85.06	54.96
60	25.69	44.35	88.32	63.64
70	24.99	54.10	87.26	56.95
80	24.33	66.06	85.87	46.86
90	21.88	70.78	85.50	42.53
100	41.38	72.29	83.13	37.63
CSSC	13.98	30.36	92.35	76.97
Our proposed method	18.51	16.93	93.42	82.27

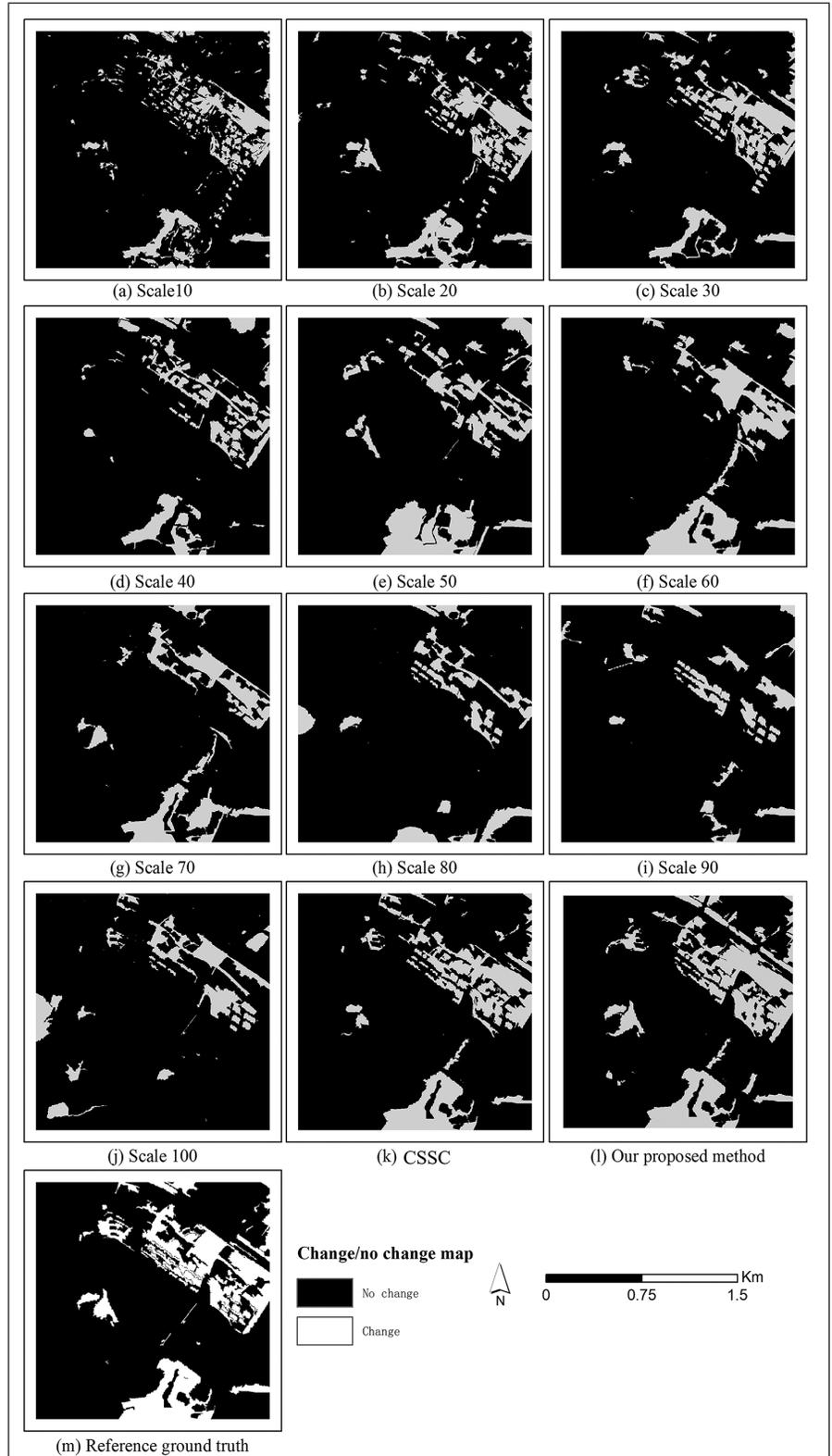


Figure 8. Binary change-detection (CD) results of data set 1 using our proposed method, single-scale, single-classifier (SSSC) methods, and the class-specific-scale, single-classifier (CSSC) method. (a to j) Binary CD results of SSSC methods at (a) scale 10, (b) scale 20, (c) scale 30, (d) scale 40, (e) scale 50, (f) scale 60, (g) scale 70, (h) scale 80, (i) scale 90, and (j) scale 100. (k) CSSC method. (l) Our proposed method. (m) Reference ground truth.

37.63%–67.90%). These results illustrate that our proposed method can significantly reduce missed detections and increase performance compared with all 10 tested SSSC methods for data set 1. The FAR of our proposed method (18.51%) is lower than those for scales 40–100 (23.38%–41.38%), illustrating that our proposed method effectively reduces false alarms compared with most of the SSSC methods.

Table 2 also compares the accuracy of our proposed method with the accuracy of the CSSC method for data set 1. The MDR of our proposed method (16.93%) is lower than for the CSSC method (30.36%), illustrating that our proposed method dramatically improves missed detections. OA (93.42%) and F1 (82.27%) for our proposed method are higher than for CSSC (OA: 92.35%; F1: 76.97%). However, the FAR of our proposed method (18.51%) is higher than for CSSC (13.98%). These results illustrate that our proposed method is not effective in reducing false alarms compared with the CSSC method.

Values for FAR, MDR, OA, and F1 of binary CD using the proposed method versus the SSSC and CSSC methods for data set 2 are shown in Table 3. The MDR of our proposed method (13.05%) is significantly lower than for all 10 SSSC scales (16.24%–54.62%). These results illustrate that our proposed method can effectively reduce missed detections compared with all 10 tested SSSC methods. Our proposed method delivers higher OA (93.18%) and F1 (83.54%) than all 10 SSSC scales (OA: 87.64%–90.78%; F1: 59.95%–76.21%), thus confirming the validity of our proposed method. However, the FAR of our proposed method is lower than for scales 20, 40, 180, and 200, illustrating that our proposed method does not effectively reduce false alarms compared with most of the tested SSSC methods.

Table 3 also compares the accuracy of our proposed method with the accuracy of the CSSC method for data set 2. The MDR of our proposed method (13.05%) is lower than for the CSSC method (19.80%), illustrating that our proposed method effectively reduces missed detections. OA (93.18%) and F1 (83.54%) for our proposed method are significantly higher than for CSSC (OA: 90.05%; F1: 76.25%), demonstrating that the class-specific classifier-selection method is much more accurate than the single-classifier approach. The FAR of our proposed method (19.61%) is lower than for CSSC (27.34%).

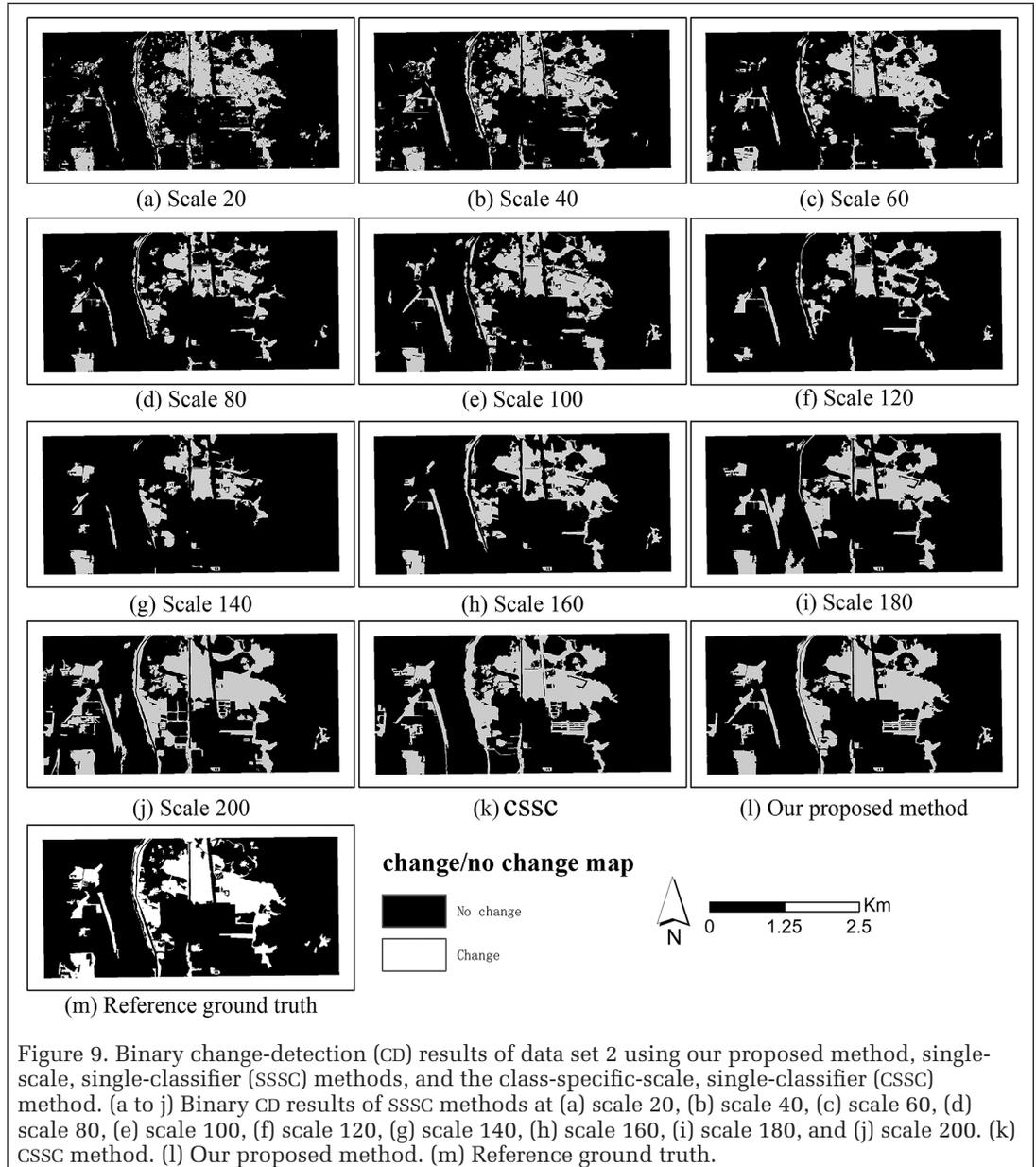


Figure 9. Binary change-detection (CD) results of data set 2 using our proposed method, single-scale, single-classifier (SSSC) methods, and the class-specific-scale, single-classifier (CSSC) method. (a to j) Binary CD results of SSSC methods at (a) scale 20, (b) scale 40, (c) scale 60, (d) scale 80, (e) scale 100, (f) scale 120, (g) scale 140, (h) scale 160, (i) scale 180, and (j) scale 200. (k) CSSC method. (l) Our proposed method. (m) Reference ground truth.

Table 3. False-alarm rate (FAR), missed-detection rate (MDR), overall accuracy (OA), and F1 values of binary change-detection results using our proposed method, single-scale, single-classifier (SSSC) methods, and the class-specific-scale, single-classifier (CSSC) method for data set 2.

Method	FAR (%)	MDR (%)	OA (%)	F1 (%)
SSSC				
20	19.82	30.97	90.44	74.19
40	19.65	31.79	90.35	73.78
60	17.73	31.54	90.78	74.73
80	14.20	48.68	88.61	64.22
100	19.45	32.35	90.31	73.54
120	15.58	53.52	87.64	59.95
140	8.07	54.62	88.33	60.77
160	16.97	36.39	90.17	72.04
180	24.63	36.28	88.63	69.06
200	30.08	16.24	89.59	76.21
CSSC	27.34	19.80	90.05	76.25
Our proposed method	19.61	13.05	93.18	83.54

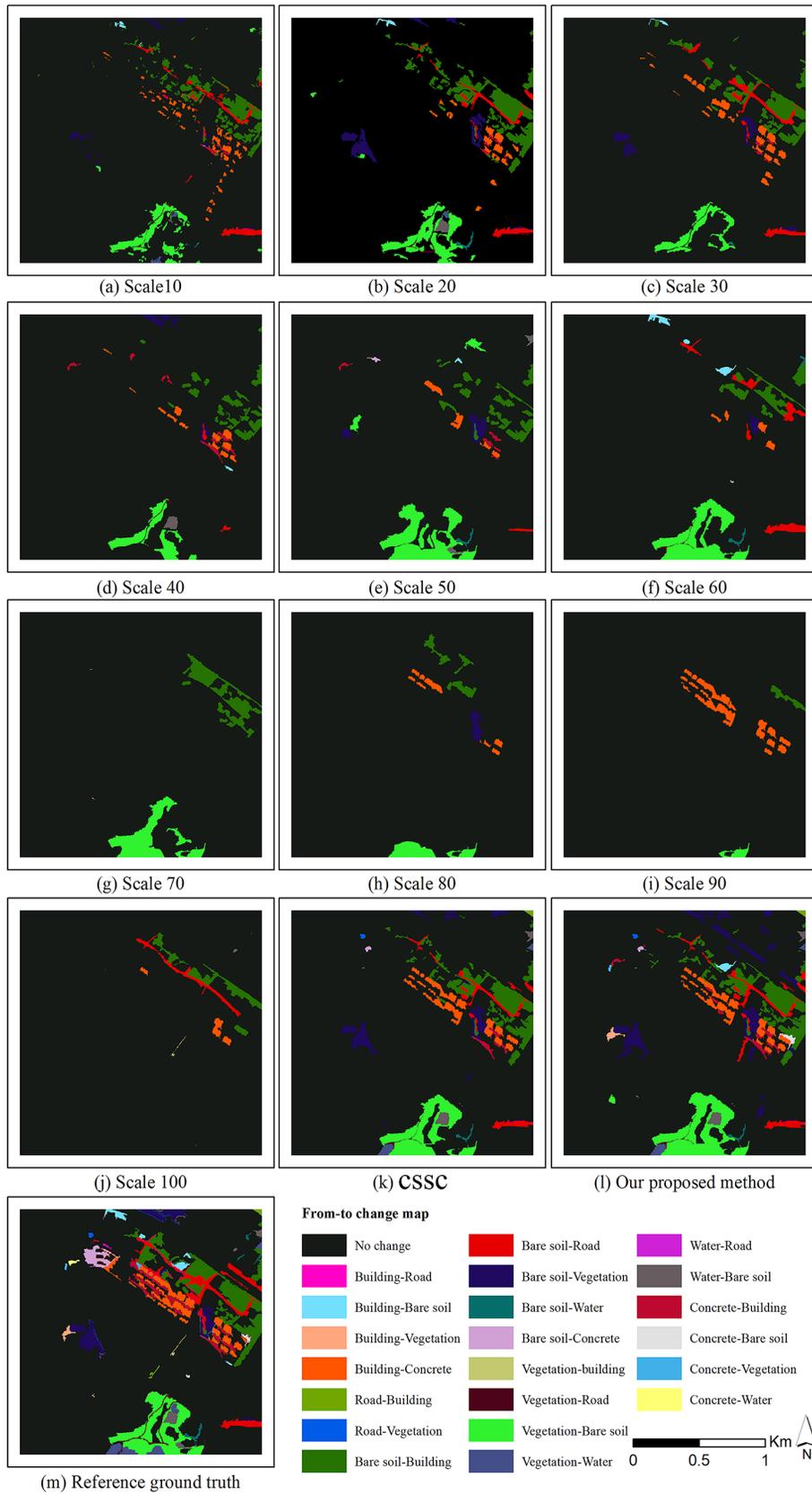


Figure 10. From-to change-detection (CD) results of data set 1 using our proposed method, single-scale, single-classifier (SSSC) methods, and the class-specific-scale, single-classifier (CSSC) method. (a to j) From-to CD results of SSSC methods at (a) scale 10, (b) scale 20, (c) scale 30, (d) scale 40, (e) scale 50, (f) scale 60, (g) scale 70, (h) scale 80, (i) scale 90, and (j) scale 100. (k) CSSC method. (l) Our proposed method. (m) Reference ground truth.

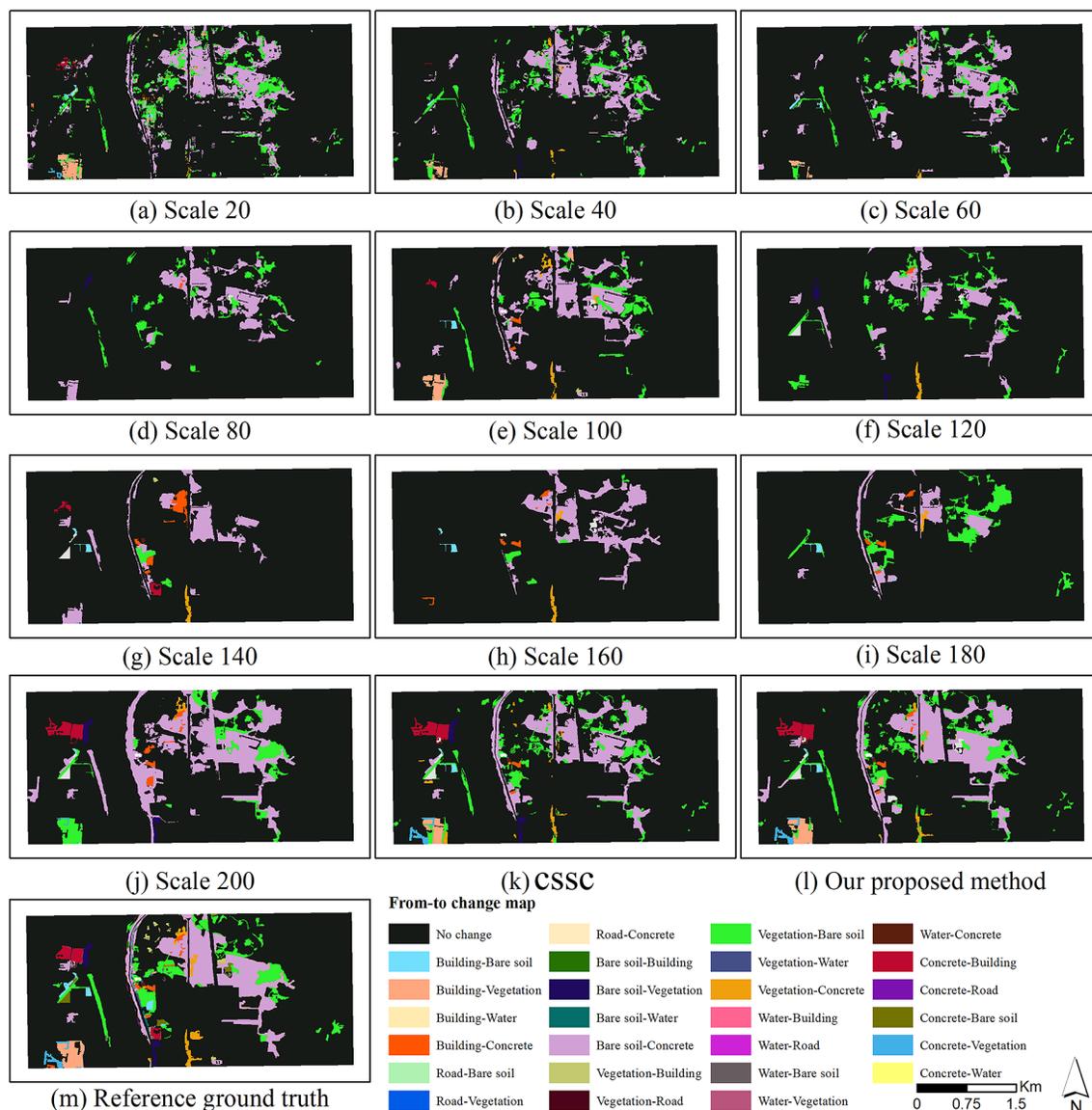


Figure 11. From-to change-detection (CD) results of data set 2 using our proposed method, single-scale, single-classifier (SSSC) methods, and the class-specific-scale, single-classifier (CSSC) method. (a to j) From-to CD results of SSSC methods at (a) scale 20, (b) scale 40, (c) scale 60, (d) scale 80, (e) scale 100, (f) scale 120, (g) scale 140, (h) scale 160, (i) scale 180, and (j) scale 200. (k) CSSC method. (l) Our proposed method. (m) Reference ground truth.

These results illustrate that our proposed method is effective in reducing false alarms compared with the CSSC method.

In order to more intuitively analyze the effectiveness of the proposed method compared with the SSSC and CSSC methods, we drew the binary and from-to CD accuracy distributions for the two data sets according to these four index values using box plots (Figure 12).

In Figure 12, the x-axis represents the accuracy indexes, and the y-axis represents their values. Blue boxes represent the SSSC method for binary CD, and pink boxes the SSSC method for from-to CD. Within these boxes, the line represents the median value of the accuracy index using SSSC methods. Blue circles represent the CSSC method for binary CD, and pink circles the CSSC method for from-to CD. Blue diamonds represent the proposed method for binary CD, and pink diamonds the proposed method for from-to CD. In the following, we present a detailed accuracy analysis of binary and from-to CD based on Figure 12.

As can be seen from the figure, our proposed method offers significantly improved performance over the tested SSSC methods and the CSSC method for both data sets in both

binary and from-to CD, as evidenced by the fact that the diamonds are above the whole box plots and circles for OA and F1 in both data sets.

Furthermore, our proposed method delivers a lower MDR than the tested SSSC methods and the CSSC method for binary and from-to CD, as the diamonds are below the whole box plots and circles for MDR in both data sets. The proposed method does not, however, have a clear advantage over the tested SSSC methods and CSSC method for binary and from-to CD when it comes to false alarms. This is because we cannot get a consistent measure of the accuracy relationship between FAR values in the two data sets. Thus the accuracy of the binary and from-to CD results at different scales fluctuates significantly, since the length of each box is very long.

#### Comparison Analyses with Other Methods

In order to assess the effectiveness of the proposed method, our binary CD results were compared with state-of-the-art techniques including PCA *k*-means (Celik 2009), OCVA (object-based change vector analysis; Sun and Chen 2010), and SVM (Cortes and Vapnik 1995). For data set 1, PCA *k*-means, OCVA

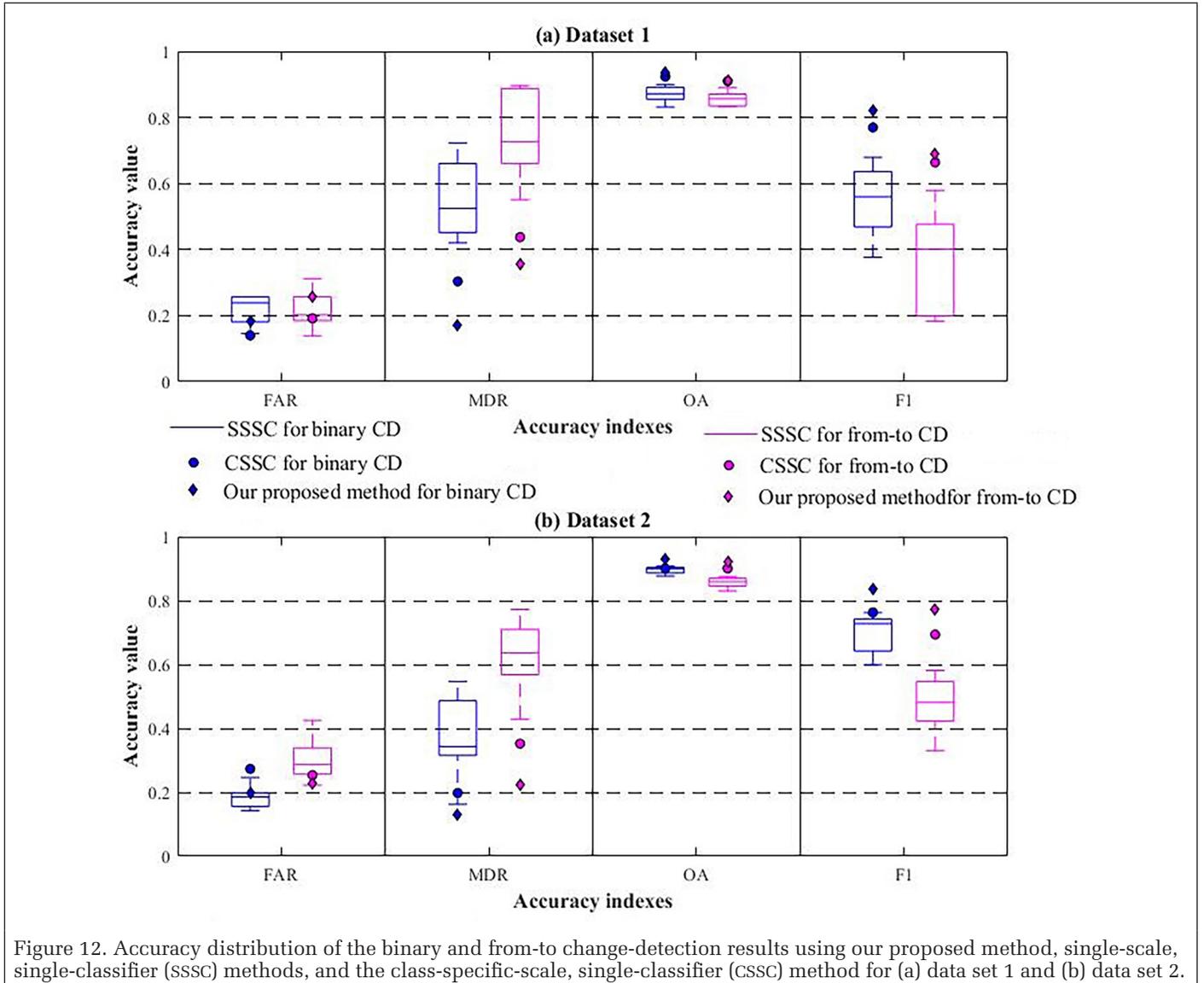


Figure 12. Accuracy distribution of the binary and from-to change-detection results using our proposed method, single-scale, single-classifier (SSSC) methods, and the class-specific-scale, single-classifier (CSSC) method for (a) data set 1 and (b) data set 2.

at single scale 100, OCVA at class-specific scales, SVM at single scale 100, and SVM at class-specific scales were conducted. For data set 2, PCA *k*-means, OCVA at single scale 200, OCVA at class-specific scales, SVM at single scale 200, and SVM at class-specific scales were conducted. The respective results are shown in Figures 13 and 14.

Visual inspection of the figures shows that our proposed method has higher accuracy on these two data sets than other methods. This can be seen by the fact that our proposed method has fewer missed detections and false alarms than

the other methods. These results indicate that our method is suitable for WorldView-2 and Gaofen-2 satellite image data. The FAR, MDR, OA, and F1 values were calculated in order to compare and validate the proposed CD method against PCA *k*-means, OCVA, and SVM. The results for the two data sets are shown in Tables 4 and 5.

As Tables 4 and 5 illustrate, our proposed method has higher OA and F1 and lower FAR and MDR than the pixel-based PCA *k*-means method in both data sets. Machine-learning methods including SVM and our proposed method have lower

Table 4. False-alarm rate (FAR), missed-detection rate (MDR), overall accuracy (OA), and F1 values of change-detection results using principal component analysis (PCA) *k*-means, object-based change vector analysis (OCVA) at scale 100, OCVA at class-specific scales, support vector machine (SVM) at scale 100, SVM at class-specific scales, and our proposed method for data set 1.

Method	FAR (%)	MDR (%)	OA (%)	F1 (%)
PCA <i>k</i> -means	65.82	64.78	76.37	34.69
OCVA at scale 100	67.62	56.96	64.63	36.96
OCVA at class-specific scales	67.94	43.42	57.18	40.93
SVM at scale 100	65.60	45.57	64.02	42.16
SVM at class-specific scales	42.90	32.74	78.17	61.76
Our proposed method	18.51	16.93%	93.42	82.27

Table 5. False-alarm rate (FAR), missed-detection rate (MDR), overall accuracy (OA), and F1 values of change-detection results using principal component analysis (PCA) *k*-means, object-based change vector analysis (OCVA) at scale 200, OCVA at class-specific scales, support vector machine (SVM) at scale 200, SVM at class-specific scales, and our proposed method for data set 2.

Method	FAR (%)	MDR (%)	OA (%)	F1 (%)
PCA <i>k</i> -means	63.55	59.95	76.48	38.17
OCVA at scale 200	87.67	68.64	65.08	17.70
OCVA at class-specific scales	88.01	69.35	65.60	17.23
SVM at scale 200	69.77	9.32	73.81	45.34
SVM at class-specific scales	48.52	29.84	88.78	59.39
Our proposed method	19.61	13.05	93.18	83.54

FAR and MDR, and higher OA and F1, than OCVA methods, illustrating that machine-learning methods deliver more accurate results.

In comparisons between single-scale methods and methods with class-specific scales, the class-specific methods perform effectively, demonstrating the validity of our proposed class-specific scale-selection method. Compared with single-classifier CD methods including SVM at scale 200 and SVM at class-specific scales, our proposed method delivers higher OA and F1 and lower FAR and MDR, demonstrating the validity of our proposed class-specific classifier-selection method.

### Discussion

In this article, we proposed a novel class-specific scale-selection method. This method uses the enhanced accuracy of an RF change detector that selects the optimal scale for each change type in land cover. So this method is not only direct and simple, it also considers scale differences between all change types in land cover. In addition, we proposed a class-specific classifier-selection method. For each change type in land cover, we use the minimum CV error in training samples as the constraint rule to select the optimal classifier from four types. Therefore, our proposed method is more precise and targeted. In applications, our proposed method is not only suitable for binary CD but also accurate and scalable for from-to CD. This method can be used for land cover map updates and is particularly well suited for urban areas.

We also recognize that there are potential limitations to this novel approach. It assumes the same classification types in multi-date images. In addition, we selected class-specific scales and class-specific classifiers for six change types in land cover, which is not refined enough for real applications. In the future, we will develop methods for selecting class-specific scales and classifiers for detecting more refined changes. Another limitation is that the proposed approach does require the availability of one reliable land cover map providing specific land cover types. Actually, there are adequate historical land cover vector maps in from the China Surveying and Mapping Department. How to effectively use historical land cover vector maps to assist CD has always been an important task in surveying and mapping. In future work, we will exploit historical multi-period land use maps for more refined CD.

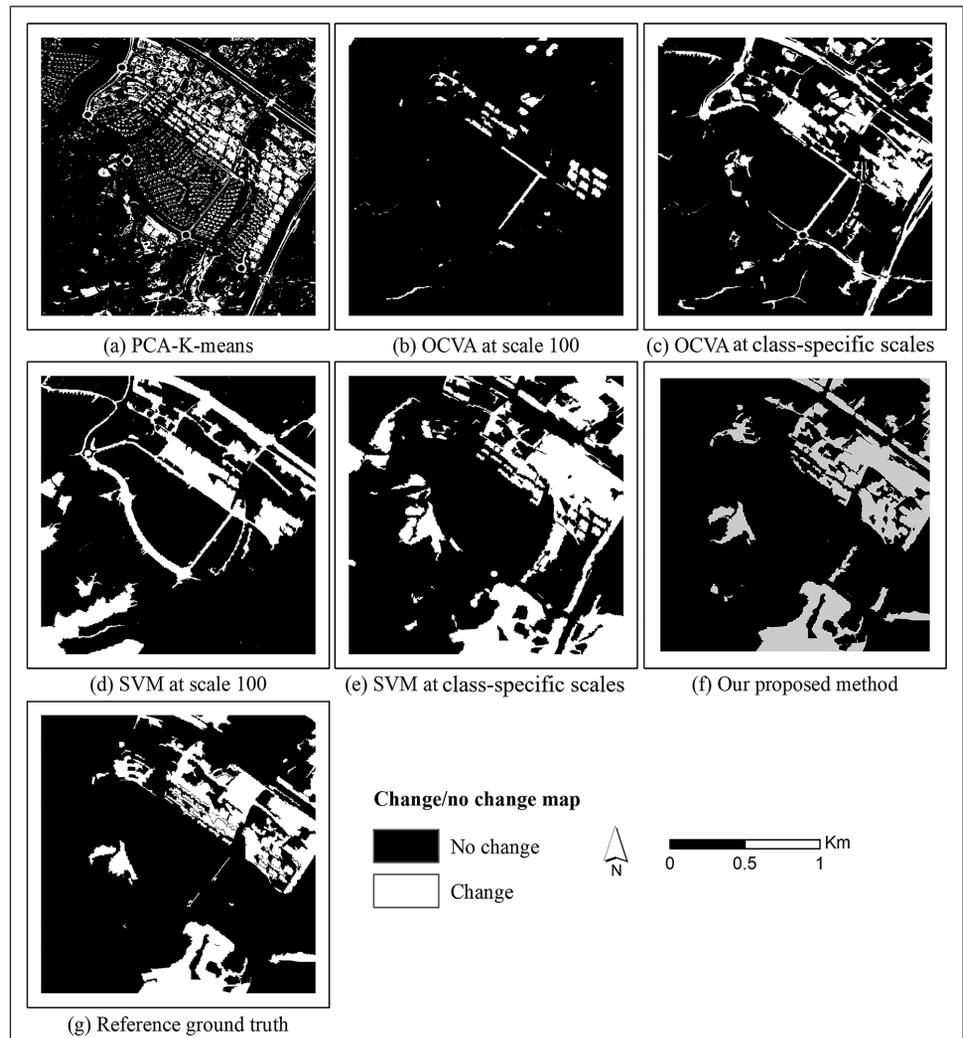


Figure 13. Binary change-detection results for data set 1 using (a) principal component analysis (PCA) k-means, (b) object-based change vector analysis (OCVA) at scale 100, (c) OCVA at class-specific scales, (d) support vector machine (SVM) at scale 100, (e) SVM at class-specific scales, (f) our proposed method, and (g) reference ground truth.

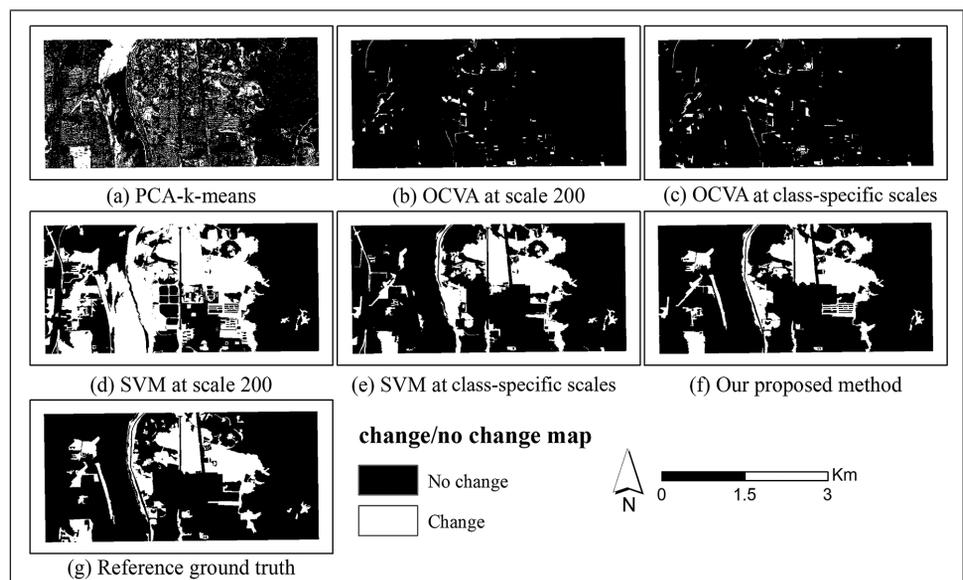


Figure 14. Binary change-detection results for data set 2 using (a) principal component analysis (PCA) k-means, (b) object-based change vector analysis (OCVA) at scale 200, (c) OCVA at class-specific scales, (d) support vector machine (SVM) at scale 200, (e) SVM at class-specific scales, (f) our proposed method, and (g) reference ground truth.

## Conclusions

In this article, we presented a novel class-specific OBCD method for urban areas. In this method, we use a class-specific scale-selection method to select the intrinsic scale for each change type in land cover. We also select a class-specific classifier for each change type in land cover. We then use the corresponding class-specific classifier to detect binary and from-to changes for each change type in land cover. In order to validate the proposed method, two data sets in urban areas were selected. Our proposed method, PCA  $k$ -means, OCVA, and SVM were used for comparative analysis. The results showed that our proposed method has higher OA and F1 and a lower MDR, proving that it is more accurate than other methods. It has the ability to deal with high levels of complexity in urban areas. In contrast to traditional methods, the proposed class-specific scale-selection method targets each change type in land cover and highlights the local changes. In addition, the proposed class-specific classifier-selection method automatically selects the corresponding classifier for each change type in land cover, which is more targeted and accurate than other approaches. Our proposed method also has some shortcomings. It needs a historical land cover vector map. In the future, we will study historical multi-period land use maps for more refined CD.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (91738301,41471354,41801344), the National Key Research and Development Program of China (grant number 2016YFB0502602), and LIESMARS Special Research Funding. The authors would like to thank the China Surveying and Mapping Department for providing the historical land cover maps, change/no-change maps, and from-to change maps. The authors declare no conflict of interest.

## References

- Anantrasirichai, N., A. Achim, J. E. Morgan, I. Erchova and L. Nicholson. 2013. SVM-based texture classification in optical coherence tomography. Pages PP–PP in *2013 IEEE 10th International Symposium on Biomedical Imaging*, held in San Francisco, Calif., 7–11 April 2013. Edited by J. Editor. City, St.: IEEE.
- Ayele, G. T., A. K. Tebeje, S. S. Demissie, M. A. Belete, M. A. Jemberrie, W. M. Teshome, D. T. Mengistu and E. Z. Teshale. 2018. Time series land cover mapping and change detection analysis using geographic information system and remote sensing, northern Ethiopia. *Air, Soil and Water Research* 11:1178622117751603.
- Bai, T., K. Sun, S. Deng, D. Li, W. Li and Y. Chen. 2018. Multi-scale hierarchical sampling change detection using Random Forest for high-resolution satellite imagery. *International Journal of Remote Sensing* 39 (21):7523–7546.
- Beitzel, S. M. 2006. *On Understanding and Classifying Web Queries* [Ph.D. thesis]. Chicago: Illinois Institute of Technology.
- Belgiu, M. and L. Drăguț. 2016. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing* 114:24–31.
- Breiman, L. 2001. Random forests. *Machine Learning* 45 (1):5–32.
- Celik, T. 2009. Unsupervised change detection in satellite images using principal component analysis and  $k$ -means clustering. *IEEE Geoscience and Remote Sensing Letters* 6 (4):772–776.
- Chen, Y., D. Ming, L. Zhao, B. Lv, K. Zhou and Y. Qing. 2018. Review on high spatial resolution remote sensing image segmentation evaluation. *Photogrammetric Engineering and Remote Sensing* 84 (10):629–646.
- Conchedda, G., L. Durieux and P. Mayaux. 2008. An object-based method for mapping and change analysis in mangrove ecosystems. *ISPRS Journal of Photogrammetry and Remote Sensing* 63 (5):578–589.
- Cortes, C. and V. Vapnik. 1995. Support-vector networks. *Machine Learning* 20 (3):273–297.
- Cristianini, N. and J. Shawe-Taylor. 2000. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge, U.K.: Cambridge University Press.
- Drăguț, L., D. Tiede and S. R. Levick. 2010. ESP: A tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data. *International Journal of Geographical Information Science* 24 (6):859–871.
- Duro, D. C., S. E. Franklin and M. G. Dubé. 2012. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sensing of Environment* 118:259–272.
- Feng, T., H. Ma, X. Cheng and H. Zhang. 2018. Calculation of the optimal segmentation scale in object-based multiresolution segmentation based on the scene complexity of high-resolution remote sensing images. *Journal of Applied Remote Sensing* 12 (2):025006.
- Feng, W., H. Sui, J. Tu, W. Huang and K. Sun. 2018a. A novel change detection approach based on visual saliency and random forest from multi-temporal high-resolution remote-sensing images. *International Journal of Remote Sensing* 39 (22):7998–8021.
- Feng, W., H. Sui, J. Tu, W. Huang, C. Xu and K. Sun. 2018b. A novel change detection approach for multi-temporal high-resolution remote sensing images based on rotation forest and coarse-to-fine uncertainty analyses. *Remote Sensing* 10 (7):1015.
- Freund, Y. and R. E. Schapire. 1997. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences* 55 (1):119–139.
- Georganos, S., T. Grippa, S. Vanhuysse, M. Lennert, M. Shimoni and E. Wolff. 2018. Very high resolution object-based land use–land cover urban classification using extreme gradient boosting. *IEEE Geoscience and Remote Sensing Letters* 15 (4):607–611.
- Gislason, P. O., J. A. Benediktsson and J. R. Sveinsson. 2006. Random forests for land cover classification. *Pattern Recognition Letters* 27 (4):294–300.
- Gong, M., J. Zhao, J. Liu, Q. Miao and L. Jiao. 2016. Change detection in synthetic aperture radar images based on deep neural networks. *IEEE Transactions on Neural Networks and Learning Systems* 27 (1):125–138.
- Grybas, H., L. Melendy and R. G. Congalton. 2017. A comparison of unsupervised segmentation parameter optimization approaches using moderate- and high-resolution imagery. *GIScience & Remote Sensing* 54 (4):515–533.
- 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., 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.
- Hussain, M., D. Chen, A. Cheng, H. Wei and D. Stanley. 2013. Change detection from remotely sensed images: From pixel-based to object-based approaches. *ISPRS Journal of Photogrammetry and Remote Sensing* 80:91–106.
- Johnson, B. and Z. Xie. 2011. Unsupervised image segmentation evaluation and refinement using a multi-scale approach. *ISPRS Journal of Photogrammetry and Remote Sensing* 66 (4):473–483.
- Jung, F. 2004. Detecting building changes from multitemporal aerial stereopairs. *ISPRS Journal of Photogrammetry and Remote Sensing* 58 (3–4):187–201.
- Lawrence, R. L., S. D. Wood and R. L. Sheley. 2006. Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (randomForest). *Remote Sensing of Environment* 100 (3):356–362.
- Li, Q., X. Huang, D. Wen and H. Liu. 2017. Integrating multiple textural features for remote sensing image change detection. *Photogrammetric Engineering and Remote Sensing* 83 (2):109–121.

- Liou, Y.-A., S. K. Kar and L. Chang. 2010. Use of high-resolution FORMOSAT-2 satellite images for post-earthquake disaster assessment: A study following the 12 May 2008 Wenchuan Earthquake. *International Journal of Remote Sensing* 31 (13):3355–3368.
- Lu, D., P. Mausel, E. Brondízio and E. Moran. 2004. Change detection techniques. *International Journal of Remote Sensing* 25 (12):2365–2401.
- Lu, M., J. Chen, H. Tang, Y. Rao, P. Yang and W. Wu. 2016. Land cover change detection by integrating object-based data blending model of Landsat and MODIS. *Remote Sensing of Environment* 184:374–386.
- Lunetta, R. S., D. M. Johnson, J. G. Lyon and J. Crotwell. 2004. Impacts of imagery temporal frequency on land-cover change detection monitoring. *Remote Sensing of Environment* 89 (4):444–454.
- Meinel, G. and M. Neubert. 2004. A comparison of segmentation programs for high resolution remote sensing data. *International Archives of Photogrammetry and Remote Sensing* 35 (Part B):1097–1105.
- Mountrakis, G., J. Im and C. Ogole. 2011. Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing* 66 (3):247–259.
- Nemmour, H. and Y. Chibani. 2006. Multiple support vector machines for land cover change detection: An application for mapping urban extensions. *ISPRS Journal of Photogrammetry and Remote Sensing* 61 (2):125–133.
- Nielsen, A. A. 2007. The regularized iteratively reweighted MAD method for change detection in multi- and hyperspectral data. *IEEE Transactions on Image Processing* 16 (2):463–478.
- Paneque-Gálvez, J., J.-F. Mas, G. Moré, J. Cristóbal, M. Orta-Martínez, A. C. Luz, M. Guèze, M. J. Macía and V. Reyes-García. 2013. Enhanced land use/cover classification of heterogeneous tropical landscapes using support vector machines and textural homogeneity. *International Journal of Applied Earth Observation and Geoinformation* 23:372–383.
- Pu, R., P. Gong, Y. Tian, X. Miao, R. I. Carruthers and G. L. Anderson. 2008. Invasive species change detection using artificial neural networks and CASI hyperspectral imagery. *Environmental Monitoring and Assessment* 140 (1–3):15–32.
- Rodríguez-Galiano, V. F., F. Abarca-Hernández, B. Ghimire, M. Chica-Olmo, P. M. Atkinson and C. Jeganathan. 2011. Incorporating spatial variability measures in land-cover classification using random forest. *Procedia Environmental Sciences* 3:44–49.
- Rokni, K., A. Ahmad, K. Solaimani and S. Hazini. 2015. A new approach for surface water change detection: Integration of pixel level image fusion and image classification techniques. *International Journal of Applied Earth Observation and Geoinformation* 34:226–234.
- Shao, Z., Y. Pan, C. Diao and J. Cai. 2019. Cloud detection in remote sensing images based on multiscale features-convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing* 57 (6):4062–4076.
- Song, Y., R. Niu, S. Xu, R. Ye, L. Peng, T. Guo, S. Li and T. Chen. 2019. Landslide susceptibility mapping based on weighted gradient boosting decision tree in Wanzhou section of the Three Gorges Reservoir Area (China). *ISPRS International Journal of Geo-Information* 8 (1):4.
- Sun, K. and Y. Chen. 2010. The application of objects change vector analysis in object-level change detection. Pages 6–7 in *Proceedings of the 3rd International Conference on Computational Intelligence and Industrial Application (PACIA 2010)*, held in Wuhan, China, November 2010. Edited by J. Editor. City, St.: Publisher.
- Tewkesbury, A. P., A. J. Comber, N. J. Tate, A. Lamb and P. F. Fisher. 2015. A critical synthesis of remotely sensed optical image change detection techniques. *Remote Sensing of Environment* 160:1–14.
- Townshend, J. R., C. O. Justice, C. Gurney and J. McManus. 1992. The impact of misregistration on change detection. *IEEE Transactions on Geoscience and Remote Sensing* 30 (5):1054–1060.
- Tu, J., D. Li, W. Feng, Q. Han and H. Sui. 2017. Detecting damaged building regions based on semantic scene change from multi-temporal high-resolution remote sensing images. *ISPRS International Journal of Geo-Information* 6 (5):131.
- Ustuner, M. and F. Balik Sanli. 2019. Polarimetric target decompositions and light gradient boosting machine for crop classification: A comparative evaluation. *ISPRS International Journal of Geo-Information* 8 (2):97.
- Verikas, A., A. Gelzinis and M. Bacauskiene. 2011. Mining data with random forests: A survey and results of new tests. *Pattern Recognition* 44 (2):330–349.
- Vinatier, F. and A. G. Arnaiz. 2018. Using high-resolution multitemporal imagery to highlight severe land management changes in Mediterranean vineyards. *Applied Geography* 90:115–122.
- Volpi, M., D. Tuia, F. Bovolo, M. Kanevski and L. Bruzzone. 2013. Supervised change detection in VHR images using contextual information and support vector machines. *International Journal of Applied Earth Observation and Geoinformation* 20:77–85.
- Walter, V. 2004. Object-based classification of remote sensing data for change detection. *ISPRS Journal of Photogrammetry and Remote Sensing* 58 (3–4):225–238.
- Wang, X., S. Liu, P. Du, H. Liang, J. Xia and Y. Li. 2018. Object-based change detection in urban areas from high spatial resolution images based on multiple features and ensemble learning. *Remote Sensing* 10 (2):276.
- Wang, Y., N. Shu, Y. Gong and X. Li. 2012. Land use change detection based on class spectral change rule. *Remote Sensing for Land & Resources* 24 (3):92–96.
- Woo, D.-M. and V. D. Do. 2015. Post-classification change detection of high resolution satellite images using AdaBoost classifier. *Advanced Science and Technology Letters* 117:34–38.
- Xun, L. and L. Wang. 2015. An object-based SVM method incorporating optimal segmentation scale estimation using Bhattacharyya Distance for mapping salt cedar (*Tamarisk* spp.) with QuickBird imagery. *GIScience & Remote Sensing* 52 (3):257–273.
- Zhang, H., J. E. Fritts and S. A. Goldman. 2008. Image segmentation evaluation: A survey of unsupervised methods. *Computer Vision and Image Understanding* 110 (2):260–280.
- Zhang, H., X. Ning, Z. Shao and H. Wang. 2019. Spatiotemporal pattern analysis of China's cities based on high-resolution imagery from 2000 to 2015. *ISPRS International Journal of Geo-Information* 8 (5):241.
- Zheng, Z., J. Cao, Z. Lv and J. A. Benediktsson. 2019. Spatial-spectral feature fusion coupled with multi-scale segmentation voting decision for detecting land cover change with VHR remote sensing images. *Remote Sensing* 11 (16):1903.

# Discovering Potential Illegal Construction Within Building Roofs from UAV Images Using Semantic Segmentation and Object-Based Change Detection

Yang Liu, Yujie Sun, Shikang Tao, Min Wang, Qian Shen, and Jiru Huang

## Abstract

A novel potential illegal construction (PIC) detection method by bitemporal unmanned aerial vehicle (UAV) image comparison (change detection) within building roof areas is proposed. In this method, roofs are first extracted from UAV images using a depth-channel improved UNet model. A two-step change detection scheme is then implemented for PIC detection. In the change detection stage, roofs with appearance, disappearance, and shape changes are first extracted by morphological analysis. Subroof primitives are then obtained by roof-constrained image segmentation within the remaining roof areas, and object-based iteratively reweighted multivariate alteration detection (IR-MAD) is implemented to extract the small PICs from the subroof primitives. The proposed method organically combines deep learning and object-based image analysis, which can identify entire roof changes and locate small object changes within the roofs. Experiments show that the proposed method has better accuracy compared with the other counterparts, including the original IR-MAD, change vector analysis, and principal components analysis-K-means.

## Introduction

Illegal construction (IC) inspection and prevention are regarded as routine tasks of urban management. ICs can be located in remote sensing scenarios by comparing the latest imagery and benchmark database containing legal constructions, such

---

Yang Liu is with Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Natural Resources, Shenzhen, 518034, China.

Min Wang is with Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Natural Resources, Shenzhen, 518034, China. Email: sysj0918@126.com.

Yang Liu, Yujie Sun, Shikang Tao, Min Wang, Qian Shen, and Jiru Huang are with Key Laboratory of Virtual Geographic Environment (Nanjing Normal University), Ministry of Education, Nanjing, 210023, China.

Yang Liu, Yujie Sun, Shikang Tao, Min Wang, Qian Shen, and Jiru Huang are with the School of Geography, Nanjing Normal University, Nanjing, 210023, China.

Yang Liu, Yujie Sun, Shikang Tao, Min Wang, Qian Shen, and Jiru Huang are with the Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing, 210023, China.

Yang Liu, Yujie Sun, Shikang Tao, Min Wang, Qian Shen, and Jiru Huang are with State Key Laboratory Cultivation Base of Geographical Environment Evolution (Jiangsu Province), Nanjing, 210023, China.

as urban development plans (Varol *et al.* 2019), cadastral data (Ostankovich and Afanasyev 2018), and city and municipality property maps (Moghadam *et al.* 2015; Khalilimoghadam *et al.* 2017), and analyzing the content differences (Ye and Chen 2015; Doxani *et al.* 2015). This kind of method can identify real ICs. However, the retrieval of legal construction data is occasionally difficult. By contrast, potential ICs (PICs) can be located by change detection techniques, which determine newly emerged buildings/construction by comparing multi-temporal imagery (Chen *et al.* 2018c; Chaabane *et al.* 2019). In such cases, the results need further validation because the illegality semantics cannot be directly obtained from imagery.

The aforementioned studies extracted the building-level, that is, the entire building appearance/disappearance, as the PICs. However, ICs may be small objects hidden within the buildings' roofs, such as photovoltaic panels, solar water heaters, and satellite television receiving antennas, which are blind corners of manual inspection. A small PIC may only occupy  $10 \times 10$  to  $20 \times 20$  in a 0.1 m aerial image and is less than one-pixel size in some space-borne images, such as *Gaofen* (GF)-1 or GF-2 images of China. Unmanned aerial vehicles (UAVs) can obtain super-high centimeter resolution images (Pajares 2015), which provide sufficient ground details and facilitate small PIC detection. PICs, including small PICs, are detected in this study using change detection techniques by comparing bitemporal UAV images.

Detecting PICs from a whole UAV image may be inefficient, and conducting roof-constrained PIC detection improves method performance. Therefore, fully convolutional networks (FCNs) (Long *et al.* 2015), which conduct pixel-level semantic segmentation, achieve good method accuracy for building (roof) extraction. Popular FCNs include UNet (Ronneberger *et al.* 2015), DeepLab (Chen *et al.* 2017a; Chen *et al.* 2017b; Chen *et al.* 2018a; Chen *et al.* 2018b), SegNet (Badrinarayanan *et al.* 2017), and PSPNet (Zhao *et al.* 2017). Among these FCN models, UNet is designed with a relatively simple network structure, low requirement for training sample volume, and good training and classification efficiency; thus, it has a good practice in remote sensing information extraction (Zhang *et al.* 2017; Li *et al.* 2018; Wang *et al.* 2020a), such as building extraction (Peng *et al.* 2019; He *et al.* 2020; Shao *et al.* 2020; Wang *et al.* 2020b). The simple network structure of the UNet also facilitates its structural modification. UNet with depth branch (UNet-DB), which integrates image spectral and depth

---

Photogrammetric Engineering & Remote Sensing  
Vol. 87, No. 4, April 2021, pp. 263–271.  
0099-1112/21/263–271

© 2021 American Society for Photogrammetry  
and Remote Sensing  
doi: 10.14358/PERS.87.4.263

features based on UNet, is proposed in this study for roof extraction. UNet-DB accurately extracts building roofs when the digital surface model (DSM) data, which offers the image depth information and can be generated by UAV oblique photogrammetry, are available.

Deep learning-based change detection generally requires large and balanced samples for model training, which is complicated in detecting PICs with small sizes, diverse types, and sparse distribution. Object-based change detection (OBCD) obtains object primitives (segments) with image segmentation and conducts unsupervised change detection on the primitives, which is suitable for small PIC detection. However, OBCD requires explicitly specified image features for change analysis. Small PICs may be low and are hard to distinguish though bitemporal DSM comparison. Image texture, edge, and other features are ineffective in change analysis due to small sizes. The image spectrum is utilized as the dominant feature for PIC detection considering the method generality. Thereafter, iteratively reweighted multivariate alteration detection (IR-MAD) (Nielsen 2007), which is robust to spectral differences between bitemporal UAV images, is selected as the change detection method. A two-step change detection strategy is designed to amend the deficiency of IR-MAD in some particular scenarios and improve the method accuracy. Through the combination of deep learning and OBCD techniques, the proposed method not only detects building-level PICs but also small PICs within the building roofs, which is an important innovation of this paper.

The rest of this paper is organized as follows. The section "Technical Route" introduces the proposed method, which includes the semantic segmentation network for roof extraction, and the two-step change detection strategy. The section "Experiments" presents two groups of comparative experiments on roof extraction and the OBCD for PIC detection. Method performance analyses and discussions are also conducted to verify the proposed method. Finally, the last section provides the conclusion.

## Technical Route

Figure 1 shows the technical route of the proposed method. Oblique photogrammetry for UAV images obtains the orthoimage and DSM imagery, which are used for roof and PIC detection. Although DSM is unnecessary, it improves roof extraction accuracy, as illustrated in the experiments. The proposed UNet-DB extracts the roofs and uses the orthoimage and DSM. If DSM is not provided, then the original UNet conducts the roof extraction without depth information.

The two-step change detection is conducted within the roofs. The building-level changes, which include building roof appearance, disappearance, and shape changes, are detected through morphological analysis on the bitemporal roofs. Image segmentation is conducted within the remaining roof areas to obtain the subroof primitives, which are subjected to OBCD for small PIC detection. The roof and subroof

primitive changes are combined and outputted as the final PIC detection results.

## Roof Extraction

Directly combining the depth and red, green, blue (RGB) images and conducting the four-band-based UNet segmentation are limited in accurate roof extraction, as shown in the experiments. Thus, UNet-DB, which constructs a separated depth branch into UNet for method improvement, is designed. The encoder of the UNet has five groups of convolution layers. Each group includes two convolution layers succeeded by one pooling layer. Such a design is followed by the depth branch. The backbone network is partially modified by adapting to the diversity of the roof shapes and sizes. A convolution layer is supplemented to the third convolution group in the UNet RGB branch to learn more low-level features than the original structure. The channel number in the fifth group is modified from 1024 to 512, thereby reducing the calculation amount.

Asymmetric convolution is further added to the network. Asymmetric convolution is generally used as an architectural element (Denton *et al.* 2014; Jaderberg *et al.* 2014) to reduce the network parameters. For example, in Inception-v3 (Szegedy *et al.* 2016), the  $7 \times 7$  convolution is replaced with  $1 \times 7$  and  $7 \times 1$  convolutions. Therefore, Ding *et al.* (2019) proposed the asymmetric convolution block (ACB), which added the outputs of the three parallel convolution branches with square, horizontal, and vertical convolution kernels (Figure

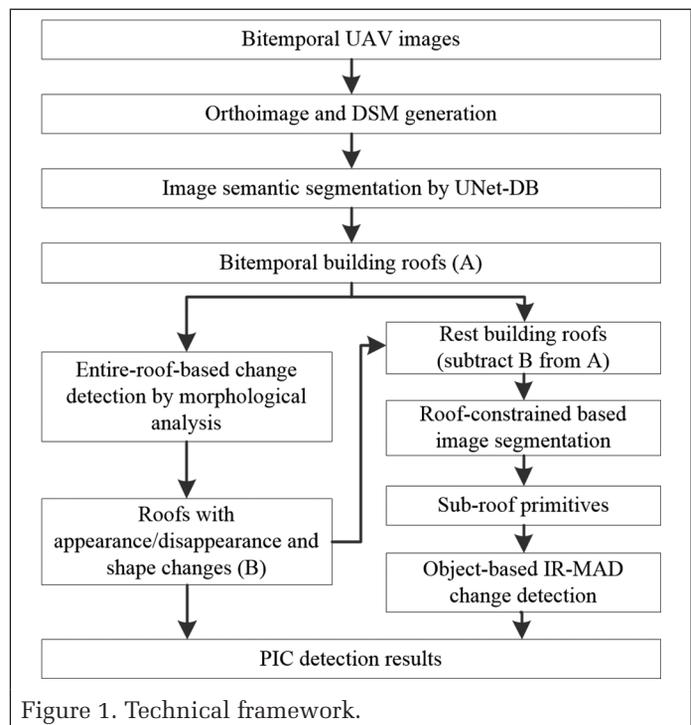


Figure 1. Technical framework.

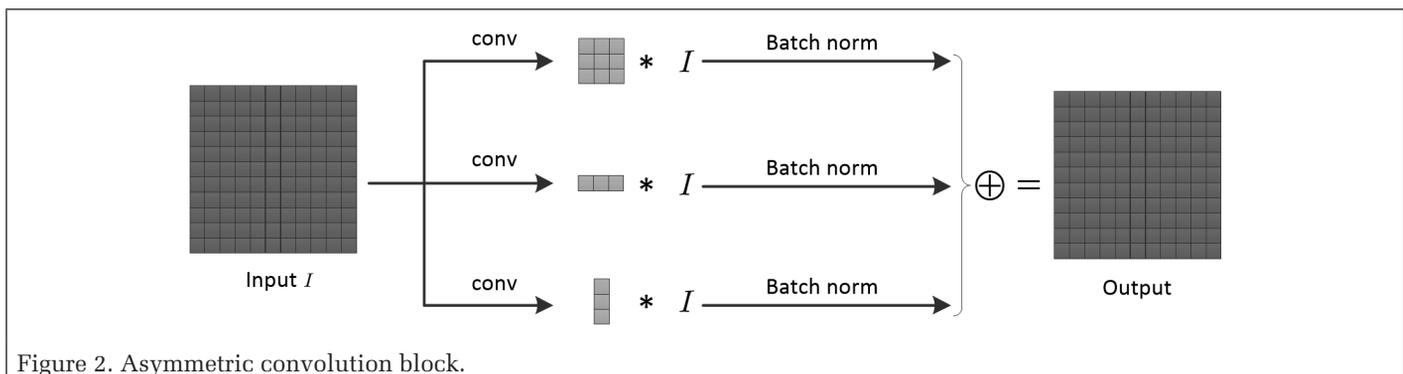
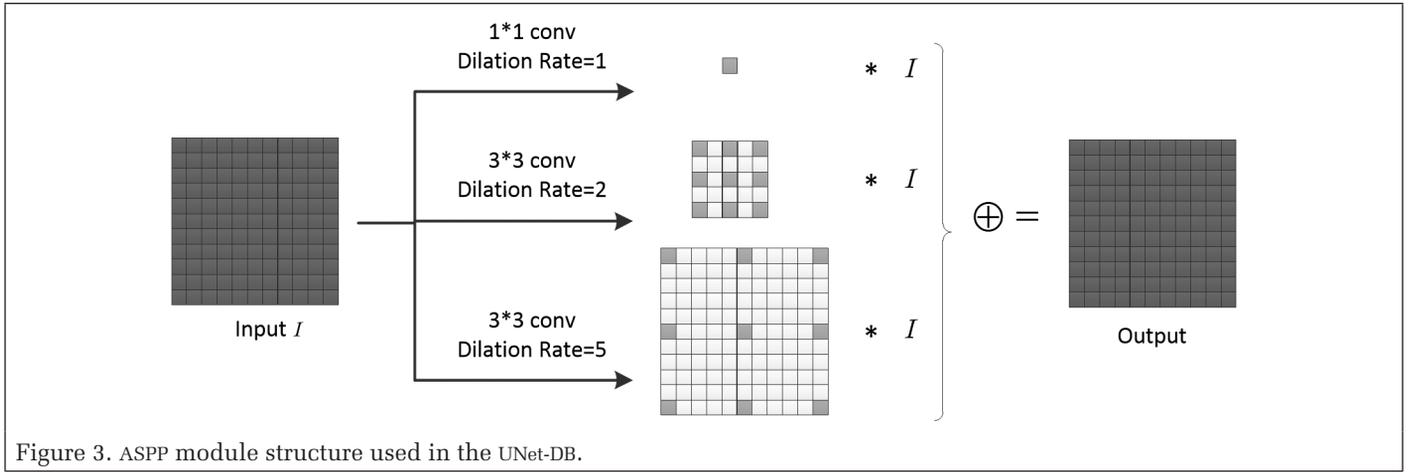


Figure 2. Asymmetric convolution block.



2). From the perspective of image filtering, ACB helps the network learn richer image features than common symmetric convolutions.

Atrous spatial pyramid pooling (ASPP) uses jumping receptive fields in image feature extraction (Yu and Koltun 2016). ASPP with different dilation rates effectively aggregates features at different scales. In this study, ASPP with dilation rates of 1, 2, and 5 is added to the network, thereby facilitating the extraction of roofs in different sizes (Figure 3).

The UNet-DB network is presented in Figure 4. The encoder includes separated depth and backbone branches, which handle the depth and RGB channels, respectively. The two branches have five groups of convolution layers. After each group of convolution, the depth branch transports the extracted features to the backbone network for information fusion. The features generated by the encoder are sent to the ASPP module for multiscale feature fusion. The decoder gradually recovers the feature maps to the original image sizes by deconvolution, conducts the prediction, and outputs the classification results.

**PIC Detection**

IR-MAD, which initially addresses the automatic radiometric calibration of multitemporal images (Nielsen *et al.* 1998), is selected as the change detection method for bitemporal UAV images with apparent radiometric differences. However, IR-MAD fails in some particular change scenarios because it only uses image spectral features. Thus, a two-step change detection scheme is designed. This scheme first detects the whole roof changes based on the roof morphology and then the subroof changes according to the reformed object-based IR-MAD (OB-IRMAD).

**Whole Roof Change Detection**

Figure 5 illustrates some roof changes in which IR-MAD fails the detection. Figure 5a shows that the newly emerged buildings and background are quite similar in color, thereby disturbing IR-MAD. Figure 5b shows the roof arrangement or orientation changes in the bitemporal images. Meanwhile, the roofs have minimal changes considering color and height. Such roof changes cannot be detected by IR-MAD or DSM

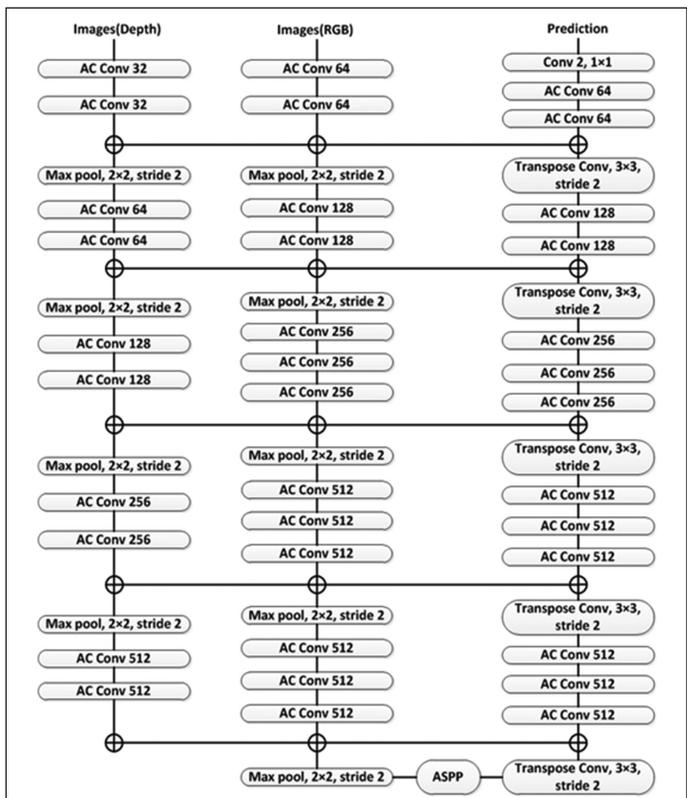
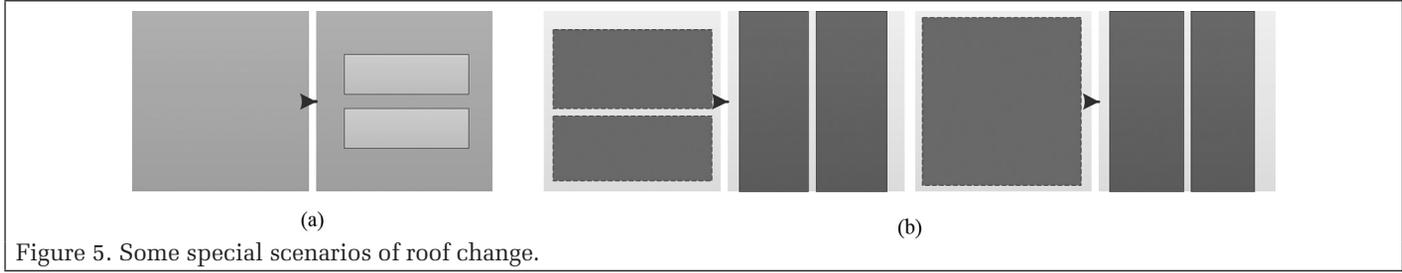


Figure 4. UNet-DB structure. “AC” means asymmetric convolution.

elevation analysis. However, the changes can be revealed by comparing the morphological difference of the bitemporal roofs. Let  $GA$  and  $GB$  be the pixel sets of the two roofs at the same location from the bitemporal images. The intersection-over-union ( $I_oU$ ) measure is calculated as follows:



$$I_{oU}(GA,GB) = \frac{GA \cap GB}{GA \cup GB}. \quad (1)$$

If the  $I_{oU}$  is less than a certain threshold, then the roofs are marked as changed and excluded from the succeeding subroof primitive-based change detection.

#### Subroof Primitive-Based Change Detection

Hard-boundary-constrained segmentation (HBC-SEG) method (Wang and Li 2014), which is an edge-embedded region-merging segmentation method, is modified and used for roof-constrained segmentation. Bitemporal images are segmented with the same inputs to obtain the consistent subroof primitives. A bidirectional matching scheme is designed to adapt to the small object-oriented change detection. Figure 6 shows that an object from one time is mapped to the same locations at the other time. The dissimilarity evaluation is conducted on the bitemporal pixels within the object. The final changes are the combination of change detection results from both directions.

Change detection is conducted by IR-MAD. Let  $X = [X_1, \dots, X_k]^T$  and  $Y = [Y_1, \dots, Y_k]^T$  ( $E\{X\} = E\{Y\} = 0$ ) be the bitemporal images, where  $k$  is the number of bands. The two vectors are defined as follows:

$$\begin{aligned} a^T X &= a_1 X_1 + \dots + a_k X_k \\ b^T Y &= b_1 Y_1 + \dots + b_k Y_k \end{aligned} \quad (2)$$

The MAD variate is defined as follows:

$$\begin{bmatrix} a_k^T X - b_k^T Y \\ \dots \\ a_1^T X - b_1^T Y \end{bmatrix}, \quad (3)$$

where  $a$  and  $b$  are the coefficients obtained from the standardized correlation analysis, which maximize the variance (i.e.,  $V\{a^T X - b^T Y\}$ ). The normalized sum of squares of the MAD variate satisfies the  $\chi^2$  distribution with  $k$  degrees of freedom:

$$T_j = \sum_{i=1}^k \left( \frac{\text{MAD}_{ij}}{\sigma_{\text{MAD}_i}} \right)^2 \in \epsilon^2(p), \quad (4)$$

where  $T_j$  is the change intensity of pixel  $j$ . The changed and unchanged pixels are distinguished in accordance with their  $T$  values.

The IR-MAD method combines the MAD and EM algorithms based on MAD and uses an iterative scheme for reasonable  $a$  and  $b$  coefficients. In principle, IR-MAD assigns high and low weights to pixels with minimal and large changes during the iterations, respectively. The weight calculation is formulated as follows:

$$w_{ij} = 1 - \frac{\int_0^{0.5T_j} t^{0.5k-1} e^{-t} dt}{\int_0^{\infty} t^{0.5k-1} e^{-t} dt}. \quad (5)$$

IR-MAD sets all pixel weights to 1.0, obtains the MAD variates, and calculates the new weights. The weights are then used in the next round of MAD variate calculation and stabilized after several rounds of iterations. Finally, K-means clustering uses the change intensity to distinguish the changed and unchanged pixels.

The pixel-based IR-MAD is easily affected by image noise and reformed to an object-based method that is suitable for the proposed task. The changed pixels are assigned to the

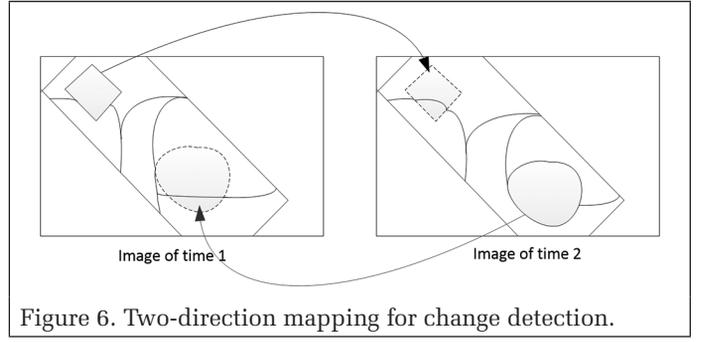


Figure 6. Two-direction mapping for change detection.

subroof primitives. A primitive is marked as changed if its changed pixels exceed 50% of its total pixel number. The reformed IR-MAD method is named OB-IRMAD in the succeeding discussion for convenience.

## Experiments

The proposed method was implemented by the research group using Visual C++ 2017 and Python 3.6 support program libraries, including GDAL 2.4, GEOS 3.5, TensorFlow 1.14.0, and OpenCV 3.0. The algorithm was tested on Windows 10, with CPU Intel i7-9800X, GPU RTX2080Ti, 64 GB of memory, and 11 GB of video memory.

### Experimental Scheme

#### Experimental Data

The UAV images captured in Yangzhou, Jiangsu Province from 2017 to 2019 were used (Figure 7). The imaging parameters are shown in Table 1. The orthoimage and DSM data were obtained through oblique photogrammetry in areas 2 and 3, and area 1 only had the orthoimages. If DSMs are available, then these images are integrated into the RGB images as the fourth band and uniformly normalized to 0–255.

Areas 1 and 2 were on the city outskirts. Area 1 included different types of apartments, villas, large sports venues, and some land under construction. Many high-rise buildings emerged in the west, which became a considerably changed subarea in the year 2019. Area 3 had a large number of low-rise buildings. A small number of changes were found by visual interpretation. However, roofs were densely distributed with complex configurations. Spectral differences between the bitemporal images were found in all these experimental areas, thereby resulting in challenging change detection.

#### Experimental Setting

UNet-DB was compared with two counterparts: UNet utilizing only RGB bands and UNet-D using RGB and depth bands for method comparison and validation. The images numbered 3, 4, 5, and 6 were selected for method comparison because they had RGB and depth bands. Approximately 20% of subareas with different types of roofs were collected from the bitemporal images, which were uniformly cropped into  $512 \times 512$  image patches. Automatic data augmentation was conducted by random image zooming, rotating, mirroring, and noise adding. The training sample set was further divided into independent training, validation, and test sets by the ratio of 5:3:2. Cross-entropy loss and RMSprop (Graves 2013) were respectively used as the loss function and the optimizer to train the FCN models. RMSprop is an improvement to the Adagrad algorithm (Duchi *et al.* 2011) with adaptive learning rates. The initial learning rate of the RMSprop optimizer was 0.001, decay was 0.9, and batch size was 20.

The extracted roofs were segmented using HBC-SEG. The scale parameter of HBC-SEG determined the average segment sizes. Scale 50 was manually tuned to obtain unbroken primitives while preventing overmerging with other objects. The



Figure 7. Bitemporal images for PIC detection. (a) Area 1, (b) Area 2, and (c) Area 3.

Table 1. Experimental data.

Experimental Area	Image ID	Imaging Time	Image Size (Pixel)	Geographic Position (WGS_84_UTM_zone_50N)	Spatial Resolution (m)	Available Data
Area 1	1	2017.04	5295 × 3945	X [721 345, 722 403]	0.2	RGB image
	2	2018.06		Y [3 585 646, 3 586 432]		
Area 2	3	2017.04	5041 × 3141	X [733 382, 734 579]	0.25	RGB Image and DSM
	4	2019.10		Y [3 588 483, 3 589 229]		
Area 3	5	2017.04	3652 × 2326	X [733 447, 734 314]	0.25	RGB Image and DSM
	6	2019.10		Y [3 587 901, 3 588 453]		

additional input setting criteria for HBC-SEG can be found in Wang and Wang (2016) and Wang *et al.* (2016, 2017a, 2017b, and 2018).

The following three classical change detection methods were selected for method comparison: change vector analysis (CVA) (Malila 1980), principal components analysis (PCA)-K-means (Celik 2009), and IR-MAD, which were originally pixel-based methods. These methods were all reformed to object-based ones for fair method comparison with OB-IRMAD. Meanwhile, CVA has no inputs. The inputs of nonoverlapping block  $H$  and dimension  $S$  were both 3.0 for PCA-K-means. The number of iterations of IR-MAD was 50. In OB-IRMAD,  $I_oU$  was 0.5, and the other settings were the same with IR-MAD.

#### Method Accuracy Evaluation

Method accuracy was evaluated on the basis of the entire imagery. Four method accuracy measures were used in the building extraction and change detection modules:

$$\begin{aligned}
 \text{Precision} &= \frac{BYN}{BYN + BNN} \\
 \text{Recall} &= \frac{BYN}{BYN + GNN} \\
 F_1 &= \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \\
 I_oU &= \frac{BYN}{BYN + BNN + GNN}
 \end{aligned} \tag{6}$$

where  $BYN$  denotes the number of real targets that are classified targets,  $BNN$  denotes the number of real nontargets but classified as targets, and  $GNN$  denotes the number of real targets but classified as nontargets. Roof extraction requires precise object boundaries; thus, target extraction accuracy was measured pixelwise. For example, measure Recall denotes the number of extracted roof pixels divided by the total number of the roof pixels. Measure  $F_1$  is the harmonic average of Precision and Recall, which evaluates the method performance from a comprehensive perspective. Measure  $I_oU$

Table 2. Method accuracy measures.

Experimental Area	Extraction Method	Experimental No.	Precision (%)	Recall (%)	$F_1$ (%)	$I_oU$ (%)	
Area2	Time 1	UNet	1	88.6	93.1	90.8	83.1
		UNet-D	2	86.4	96.2	91.0	83.6
		UNet-DB	3	97.8	96.5	97.1	94.4
	Time 2	UNet	4	96.2	97.2	96.7	93.7
		UNet-D	5	91.7	97.5	94.5	97.5
		UNet-DB	6	97.5	97.1	97.3	94.7
Area3	Time 1	UNet	7	68.6	94.7	79.5	66.0
		UNet-D	8	82.4	97.9	89.5	80.9
		UNet-DB	9	94.2	97.0	95.6	91.5
	Time 2	UNet	10	75.6	95.4	84.3	72.9
		UNet-D	11	92.6	97.5	95.0	90.5
		UNet-DB	12	94.0	97.4	95.7	91.8

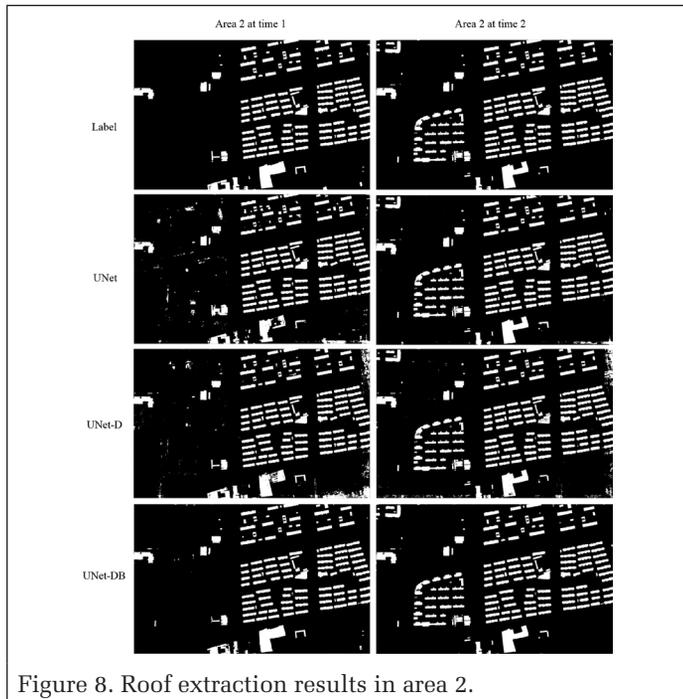


Figure 8. Roof extraction results in area 2.

denotes the set intersection divided by the set union of the extracted and real roof pixels.

The change detection accuracy was measured objectwise. Accordingly, except for  $I_oU$ , Precision, Recall, and  $F_1$  were selected as the accuracy measures. For example, Precision denotes the extracted real PIC number divided by the real PIC number, and the other measures similarly function. All the real PICs were obtained from the bitemporal subroof primitives with pixels exceeding 50%, thereby forming the reference maps.

### Performance Analysis

#### Roof Extraction

Qualitative and quantitative analyses were conducted to validate the proposed roof extraction scheme in areas 2 and 3. The results are shown in Table 2, Figures 8–10.

**Quantitative analysis:** UNet Recall exceeded 93% on all the four images, whereas Precision and  $I_oU$  significantly fluctuated (i.e., the lowest rate is 68% on experiment no. 7), thereby indicating that UNet had more false alarms than the other methods. Although with some exceptions on experimental nos. 1 and 4, UNet-D performed better than UNet on these accuracy measures in total. UNet-D increased more than 10% considering Precision,  $F_1$ , and  $I_oU$  on experimental nos. 7–8 compared with UNet. Additional depth information was effective in improving roof extraction accuracy.

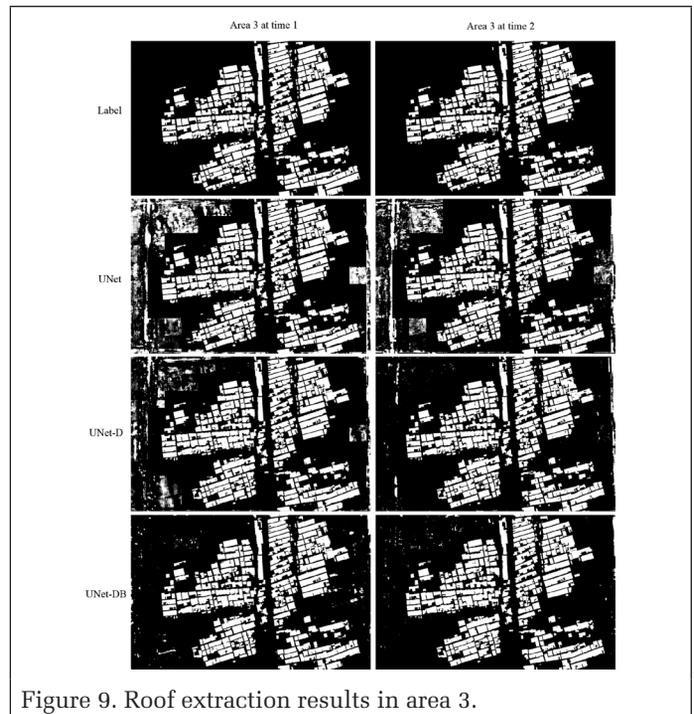


Figure 9. Roof extraction results in area 3.

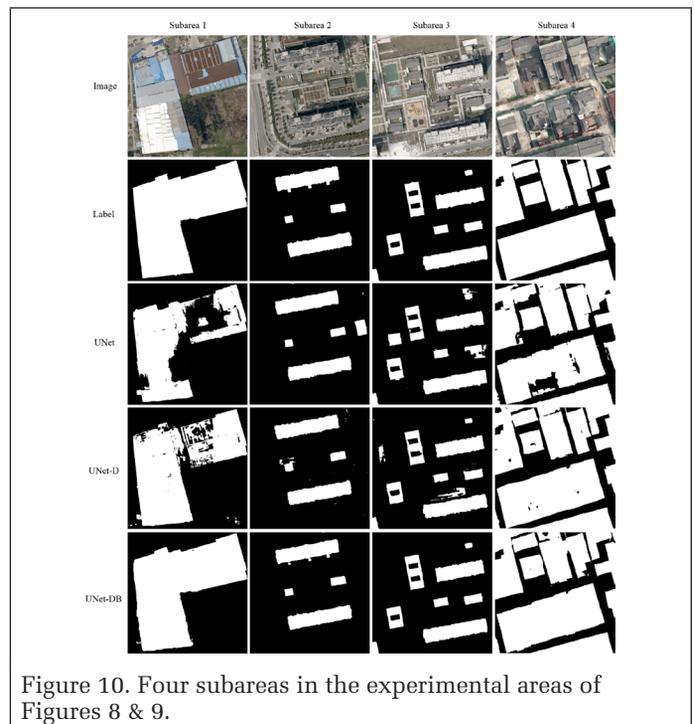


Figure 10. Four subareas in the experimental areas of Figures 8 & 9.

UNet-DB was significantly better than UNet-D considering Precision, which exceeded 94% in all experimental areas. Considering Recall, UNet-DB slightly performed weaker than UNet-D overall. Accordingly, UNet-DB was better than UNet-D on comprehensive measure  $F_1$  on all the experiments. UNet-DB also completely outperformed UNet-D on  $I_oU$ . Such evidence indicated that the depth utilization and network reform schemes of UNet-DB further improved the method performance compared with those of UNet-D.

**Qualitative analysis:** UNet extracted most roofs mixed with a large portion of bare land, vegetation, and roads. UNet-D and UNet-DB reduced such overextractions (Figure 9). Figure 10 shows some subareas in Figures 8–9. Subarea 1 had large pieces of different-colored factory roofs. UNet generally

performed in this area. UNet-D performed better than UNet

Table 3. Accuracy measures of the change detection methods.

Experimental Area	Accuracy Measure	CVA (%)	PCA-K-Means (%)	IR-MAD (%)	OB-IRMAD (%)
Area 1	Precision	29.52	25.33	51.36	51.54
	Recall	60.66	61.48	74.59	82.79
	F1	39.71	35.88	60.83	63.53
Area 2	Precision	29.92	21.50	53.93	55.78
	Recall	77.59	66.38	67.24	95.69
	F1	43.19	32.48	59.85	70.48
Area 3	Precision	26.50	22.95	39.30	39.30
	Recall	76.76	66.90	76.06	76.06
	F1	39.40	34.18	51.82	51.82

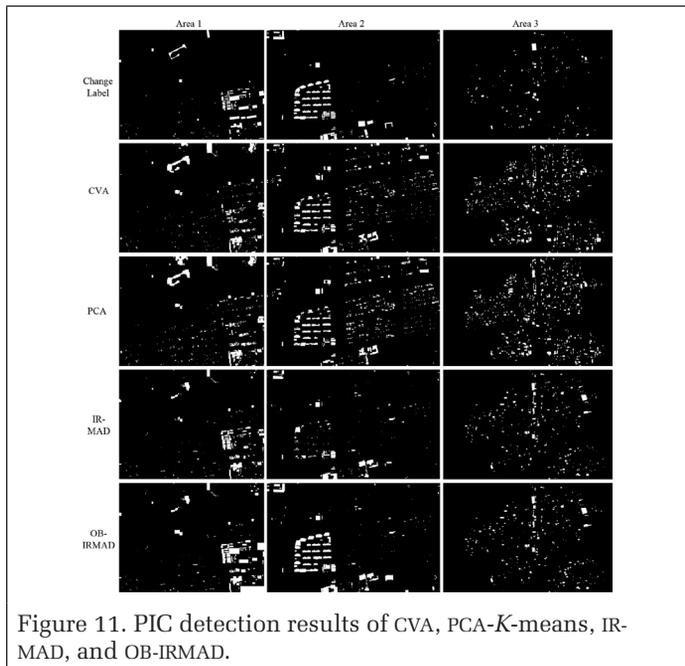


Figure 11. PIC detection results of CVA, PCA-K-means, IR-MAD, and OB-IRMAD.

but still demonstrated defects (i.e., holes), whereas UNet-DB obtained the most complete roofs. In subareas 2 and 3, some playgrounds and parking lots were extracted as roofs by UNet. Some fences were also extracted as roofs by UNet-D. UNet-DB optimally performed with small errors around the roof boundaries. In subarea 4, UNet extracted roofs but with many holes, and UNet-D eliminated some of these roofs. UNet-DB further eliminated these holes and achieved optimal results. Dilated convolutions in UNet-DB expanded the receptive field, thereby facilitating the extraction of roofs in different sizes. Asymmetric convolution learned directional features, which adapted to roofs in different orientations and shapes.

#### PIC Detection

**Quantitative analysis:** The statistics on the change detection accuracy are listed in Table 3. IR-MAD and CVA had similar Recall statistics, whereas IR-MAD Precision exceeded 15% over CVA and PCA-K-means. Accordingly, IR-MAD had superior  $F_1$  statistics to CVA and PCA-K-means. OB-IRMAD further improved the method performance compared with IR-MAD and was better than all the other methods in the Precision, Recall, and  $F_1$  measures.

**Qualitative analysis:** Figure 11 shows the change detection results in the three areas. PCA-K-means and CVA had evident false alarms. IR-MAD significantly reduced such errors but lost many changes in the lower right corner of area 1 and the lower left corners of area 2. These changed areas had similar tones to the background. OB-IRMAD recovered these changes through morphological analysis (fifth row in Figure 11).

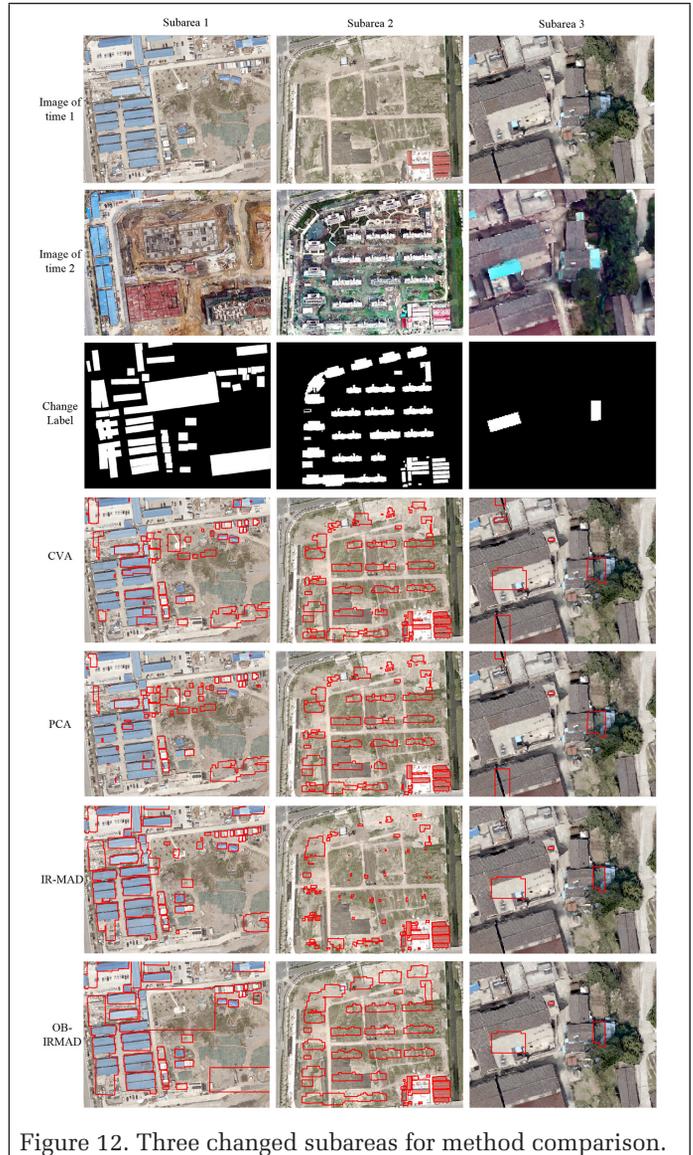


Figure 12. Three changed subareas for method comparison.

Figure 12 shows three subareas for detailed method comparison. Several buildings emerged/disappeared within the construction sites in subareas 1 and 2. PCA-K-means and CVA had missed detections in subarea 1, typically in the blue roof areas. IR-MAD had missed detection in subarea 2 due to the tone similarity of the lost roofs to the background in the first time image. In the two subareas, OB-IRMAD successfully extracted the above roof changes by utilizing the morphological and spectral features. Subarea 3 had many low-rise buildings. The IR-MAD and OB-IRMAD correctly detected the two change areas. By contrast, PCA-K-means and CVA had extra false detections (e.g., shadows between roofs). Overall, IR-MAD was superior to CVA and PCA-K-means based on the aforementioned accuracy analyses. However, the spectral feature-based IR-MAD occasionally omitted some changes. OB-IRMAD remedied such defects and obtained optimal performance by adding the morphological analysis.

#### Influence of Segmentation Scale

An appropriate segmentation scale is important for small PIC detection. Many factors, including image resolution, image complexity, amount of noise, and application goals, require consideration. The scale was set to 50, which was a generic value in the experiments. Figure 13 shows the segmentation results with a scale varying from 30 to 70. At scale 30, the sub-roof primitives were fragmentary, and the over-segmentation

errors were evident. Many under-segmentation errors emerged at scale 70 due to the merging of subroof primitives with the surrounding objects. Accordingly, scale 50 was tuned as the balanced input and generically used in all the experiments.

#### Discussions

**Building extraction network:** The UNet network, which was used in this study for building roof extraction, was also the backbone network for depth information fusion because of its simplicity, low requirement of training samples, and good method accuracy on high-spatial resolution images. However, numerous advanced models, which have potentials in accurate building extraction, have emerged (Ji *et al.* 2018; Hu *et al.* 2019; Shao *et al.* 2019; Ahmed *et al.* 2020; Zhang *et al.* 2020a; Zhang *et al.* 2020b). Designing additional sophisticated networks will be attempted by combining the depth branch fusion scheme with other network structures/substructures for building extraction.

**Utilization of the depth information:** Depth information was integrated into the semantic segmentation to improve the roof extraction accuracy. Moreover, depth information facilitates roof sample collection for network training. Numerous building roofs can be automatically extracted by simple knowledge rules involving image spectral and depth information, which reduces the labor intensity. By contrast, the depth information was not used in the change detection stage because small PICs may be low in height. However, the whole roof and tall PIC changes can be detected via bitemporal DSM elevation analysis. Depth information will also be comprehensively engaged in the OBCD for further method accuracy improvement in future studies.

**PIC detection based on deep learning:** IR-MAD only utilized image spectral features in change detection and unsatisfactorily performed in some complex scenarios (e.g., experimental area 3 of this study). Deep learning-based change detection methods are currently popular research topics. However, training sample collection and model training are difficult due to the small size, diverse type, and sparse distribution of the PICs. The knowledge transfer capability of current deep learning-based change detection methods is limited when the sample size is insufficient. Different regions often need incremental training, which is time consuming and laborious considering engineering applications. With the development of automatic sample augmentation techniques, Siamese network-based methods (Zagoruyko and Komodakis 2015) will be investigated for accurate and robust change detection.

#### Conclusion

This study proposed a novel PIC detection method based on UAV imagery. The proposed method involves roof extraction by semantic segmentation, roof segmentation, and subroof primitive change detection combining morphological and spectral features. The proposed method detects newly emerged or disappeared roofs and small PICs within the roofs through the combination of deep learning technology (building semantic segmentation) and object-based image analysis (roof segmentation and change detection). Further research will investigate the comprehensive utilization of the DSM data in change detection and the adaptation of the deep learning-based change detection method to small objects.

#### Acknowledgment

This work is jointly supported by the National Natural Science Foundation of China [42071301, 41671341], the National Key Research and Development Program of China [2017YFB0503902], the Major Science and Technology Program for Water Pollution Control and Treatment [2017ZX07302003], and the Project Supported by the

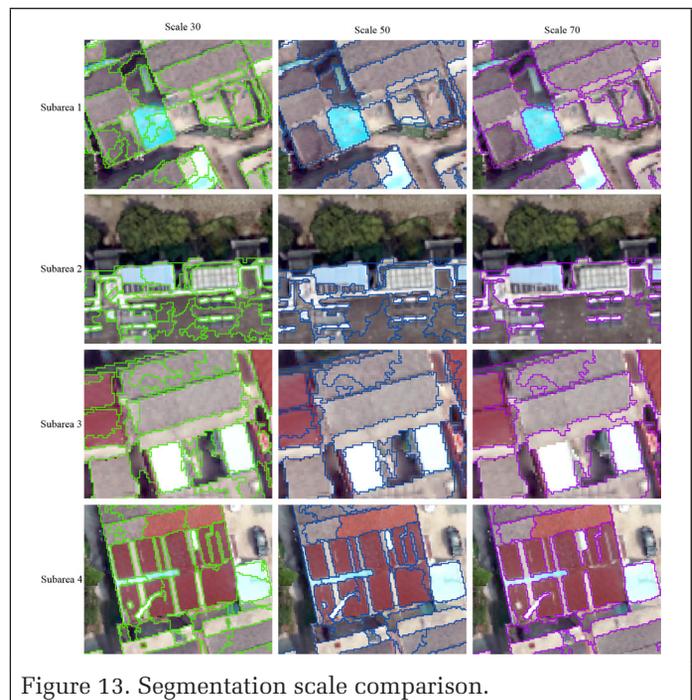


Figure 13. Segmentation scale comparison.

Open Fund of Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Natural Resources [KF-2019-04-008].

#### References

- Ahmed, N., R. B. Mahbub and R. M. Rahman. 2020. Learning to extract buildings from ultra-high-resolution drone images and noisy labels. *International Journal of Remote Sensing*:1–22.
- Badrinarayanan, V., A. Kendall and R. Cipolla. 2017. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 39 (12):2481–2495.
- Celik, T. 2009. Unsupervised change detection in satellite images using principal component analysis and k-means clustering. *IEEE Geoscience and Remote Sensing Letters* 6 (4):772–776.
- Chaabane, F., S. Réjichi, C. Kefi, H. Ismail and F. Tupin. 2019. VHR satellite image time series analysis for illegal building monitoring using multi-dimensional histogram earth mover's distance. Pages 9398–9401 in *Proceedings 2019 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2019)*, held in Yokohama, Japan.
- Chen, L. C., G. Papandreou and I. Kokkinos. 2017a. DeepLab: Semantic image segmentation with deep convolutional nets Atrous convolution and fully connected CRFs. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 40 (4):834–848.
- Chen, L. C., G. Papandreou, I. Kokkinos, K. Murphy and A. Yuille. 2018a. Deeplab: Semantic image segmentation with deep convolutional nets, Atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40 (4):834–848.
- Chen, L. C., G. Papandreou, F. Schroff and H. Adam. 2017b. Rethinking Atrous convolution for semantic image segmentation. <<https://arxiv.org/abs/1706.05587>>.
- Chen, L. C., Y. Zhu, G. Papandreou, F. Schroff and H. Adam. 2018b. Encoder-decoder with Atrous separable convolution for semantic image segmentation. Pages 833–851 in *Proceedings European Conference on Computer Vision (ECCV)*, held in Munich, Germany.
- Chen, X. F., X. G. Zhang, R. K. He and Y. Wang. 2018c. An algorithm to detect illegal buildings using color transferring and texture difference. Pages 545–550 in *Proceedings 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS)*, held in Beijing, China.

- Denton, E. L., W. Zaremba, J. Bruna, Y. Lecun and R. Fergus. 2014. Exploiting linear structure within convolutional networks for efficient evaluation. <<https://arxiv.org/abs/1404.0736>>.
- Ding, X., Y. Guo, G. Ding and J. Han. 2019. Acnet: Strengthening the kernel skeletons for powerful CNN via asymmetric convolution blocks. Pages 1911–1920 in *Proceedings IEEE/CVF International Conference on Computer Vision (ICCV)*, held in Seoul, Korea.
- Doxani, G., K. Karantzas and M. Tsakiri-Strati. 2015. Object-based building change detection from a single multispectral image and pre-existing geospatial information. *Photogrammetric Engineering & Remote Sensing* 81:481–489.
- Duchi, J., E. Hazan and Y. Singer. 2011. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research* 12:257–269.
- Graves, A. 2013. Generating sequences with recurrent neural networks. <<https://arxiv.org/abs/1308.0850>>.
- He, N., L. Fang and A. J. Plaza. 2020. Hybrid first and second order attention UNet for building segmentation in remote sensing images. *Science China (Information Sciences)* 63 (04):69–80.
- Hu, H., S. Cai, W. Wang, P. Zhang and Z. Li. 2019. A semantic segmentation approach based on DeepLab network in high-resolution remote sensing images. Pages 292–304 in *Proceedings International Conference on Image and Graphics*, held in Beijing, China.
- Jaderberg, M., A. Vedaldi and A. Zisserman. 2014. Speeding up convolutional neural networks with low rank expansions. *Computer Science* 4 (4):XIII.
- Ji, S., S. Wei and M. Lu. 2018. A scale robust convolutional neural network for automatic building extraction from aerial and satellite imagery. *International Journal of Remote Sensing*:1–15.
- Khalilimoghadam, N., M. R. Delavar and P. Hanachi. 2017. Performance evaluation of three different high resolution satellite images in semi-automatic urban illegal building detection. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*:505–514.
- Li, R., W. Liu, L. Yang, Y. Lei, S. Shihao, H. Wei, Z. Fan and L. Wei. 2018. DeepUNet: A deep fully convolutional network for pixel-level sea-land segmentation. *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing* 11 (11):3954–3962.
- Long, J., E. Shelhamer and T. Darrell. 2015. Fully convolutional networks for semantic segmentation. Pages 3431–3440 in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, held in Boston, Mass.
- Malila, W. A. 1980. Change vector analysis: An approach for detecting forest changes with Landsat. Pages 326–335 in *Proceedings of the 1980 Machine Processing of Remotely Sensed Data Symposium*, held in West Lafayette, Ind.
- Moghadam, N. K., M. R. Delavar and P. Hanachee. 2015. Automatic urban illegal building detection using multi-temporal satellite images and geospatial information systems. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 40 (1):387–393.
- Nielsen, A. 2007. The regularized iteratively reweighted MAD method for change detection in multi- and hyperspectral data. *IEEE Transactions on Image Processing* 16:463–478.
- Nielsen, A., K. Conradsen and J. Simpon. 1998. Multivariate alteration detection (MAD) and MAF postprocessing in multispectral, bitemporal image data: New approaches to change detection studies. *Remote Sensing Environment* 64:1–19.
- Ostankovich, V. and I. Afanasyev. 2018. Illegal buildings detection from satellite images using googlenet and cadastral map. Pages 616–623 in *Proceedings 2018 IEEE 9th International Conference on Intelligent Systems (IS)*, held in Madeira Island, Portugal.
- Pajares, G. 2015. Overview and current status of remote sensing applications based on unmanned aerial vehicles (UAVs). *Photogrammetric Engineering & Remote Sensing* 81:281–329.
- Peng, D., Y. Zhang and H. Guan. 2019. End-to-end change detection for high resolution satellite images using improved UNet++. *Remote Sensing* 11 (11):1382–1405.
- Ronneberger, O., P. Fischer and T. Brox. 2015. U-Net: Convolutional networks for biomedical image segmentation. Pages 234–241 in *Proceedings International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, held in Munich, Germany.
- Shao, Z., Y. Pan, C. Diao and J. Cai. 2019. Cloud detection in remote sensing images based on multiscale features-convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing* 57 (6):4062–4076.
- Shao, Z., P. Tang, Z. Wang, N. Saleem, S. Yam and C. Sommai. 2020. BRNNet: A fully convolutional neural network for automatic building extraction from high-resolution remote sensing images. *Remote Sensing* 12 (6):1050–1067.
- Szegedy, C., V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna. 2016. Rethinking the inception architecture for computer vision. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*:2818–2826.
- Varol, B., E. Ö. Yilmaz, D. Maktav, S. Bayburt and S. Gürdal. 2019. Detection of illegal constructions in urban cities: Comparing LIDAR data and stereo KOMPSAT-3 images with development plans. *European Journal of Remote Sensing* 52 (1):335–344.
- Wang, L., X. H. Ma and Y. Ye. 2020a. Computer vision-based road crack detection using an improved I-UNet convolutional networks. Pages 539–543 in *Proceedings Chinese Control and Decision Conference*.
- Wang, M., Q. Cui, Y. Sun and Q. Wang. 2018. Photovoltaic panel extraction from very high-resolution aerial imagery using region-line primitive association analysis and template matching. *ISPRS Journal of Photogrammetry and Remote Sensing* 141:100–111.
- Wang, M., J. Huang and D. Ming. 2017a. Region-line association constraints for high-resolution image segmentation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10 (2):628–637.
- Wang, M. and R. Li. 2014. Segmentation of high spatial resolution remote sensing imagery based on hard-boundary constraint and two-stage merging. *IEEE Transactions on Geoscience and Remote Sensing* 52:5712–5725.
- Wang, M., J. Xing, J. Wang and G. Lv. 2017b. Technical design and system implementation of region-line primitive association framework. *ISPRS Journal of Photogrammetry and Remote Sensing* 130:202–216.
- Wang, M. and J. Wang. 2016. A region-line primitive association framework for object-based remote sensing image analysis. *Photogrammetric Engineering and Remote Sensing* 82 (2):149–159.
- Wang, S., X. Hou and X. Zhao. 2020b. Automatic building extraction from high-resolution aerial imagery via fully convolutional encoder-decoder network with non-local block. *IEEE Access* 8:7313–7322.
- Ye, S. and D. Chen. 2015. An unsupervised urban change detection procedure by using luminance and saturation for multispectral remotely sensed images. *Photogrammetric Engineering & Remote Sensing* 81 (8):637–645.
- Yu, F. and V. Koltun. 2016. Multi-scale context aggregation by dilated convolutions. *IEEE Transactions on Computer Vision and Pattern Recognition*. <<https://arxiv.org/abs/1511.07122>>.
- Zagoruyko, S. and N. Komodakis. 2015. Learning to compare image patches via convolutional neural networks. Pages 4353–4361 in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Zhang, R., Z. Shao, X. Huang, J. Wang and D. Li. 2020a. Object detection in UAV images via global density fused convolutional network. *Remote Sensing* 12 (19):3140–3157.
- Zhang, Y., W. Li, W. Gong, Z. Wang and J. Sun. 2020b. An improved boundary-aware perceptual loss for building extraction from VHR images. *Remote Sensing* 12 (7):1195–1224.
- Zhang, Z., Q. Liu and Y. Wang. 2017. Road extraction by deep residual U-Net. *IEEE Geoscience & Remote Sensing Letters* 99:1–5.
- Zhao, H., J. Shi, X. Qi, X. Wang and J. Jia. 2017. Pyramid scene parsing network. Pages 6230–6239 in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, held in Honolulu, Hawaii.

# Call for Submissions

## Remote Sensing Monitoring for Urban Environment

*Photogrammetric Engineering and Remote Sensing (PE&RS)* is seeking submissions for a special issue on Remote Sensing Monitoring for Urban Environment.

Urban remote sensing provides images with multiple spatio-temporal-spectral attributes, which can provide qualitative, quantitative, dynamic and comprehensive information and support for urban environmental monitoring and evaluation, and serve urban planning and management, ecological environment protection. In recent decades, global urban areas have been rapidly expanding, especially in developing countries. Rapid urbanization, along with manufacturing industries and large number of vehicles has resulted in serious environmental problems, called “urban diseases”, including increased vulnerability to natural hazards, natural vegetation cover decline and arable land loss, urban heat islands, air pollution, hydrological circle alteration and biotic homogenization. Urban ecosystems are strongly influenced by anthropogenic activities. Considering this, this special issue of PE&RS is aimed at reporting novel studies on exploiting remote sensing big data to monitor and improve urban environment, and showing the potential of remote sensing in developing sustainable cities, including but not limited to:

- Urban thermal-environment remote sensing
- Remote sensing image acquisition and processing for urban environment
- Remote sensing dynamic monitoring of urban expansion
- Remote sensing change detection of urbanization
- Remote sensing retrieval of urban ecological environment
- Remote sensing evaluation of urban human settlements
- Urban sustainability indicators and assessment
- Urban environmental monitoring

Papers must be original contributions, not previously published or submitted to other journals. Submissions based on previous published or submitted conference papers may be considered provided they are considerably improved and extended. Papers must follow the instructions for authors at <http://asprs-pers.edmgr.com/>.

### Important Dates (Tentative)

- March 1, 2021—Submission system opening
- September 31, 2021—Submission system closing
- Planned publication date is December 2021
- Submit your manuscript to <http://asprs-pers.edmgr.com/> by September 31, 2021.

### Guest Editors

#### **Zhenfeng Shao, Wuhan University, China**

Prof. Zhenfeng Shao, Professor at the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, China. His research interests include urban remote sensing. He is now an associate editor of Email: [shaozhenfeng@whu.edu.cn](mailto:shaozhenfeng@whu.edu.cn).

#### **Orhan Altan, Istanbul Technical University, Turkey**

Prof. Orhan ALTAN, Professor at the Department of Geomatics Engineering, Istanbul Technical University, Turkey. He is Past President and Honorary Member of ISPRS, Honorary Member of Science Academy. He has published more than 200 scientific papers in scientific journals and conferences, and editor or co-editor of more than 20 international books. Email: [oaltan@itu.edu.tr](mailto:oaltan@itu.edu.tr).

#### **J.L.van Genderen, University of Twente, Netherlands**

Professor J.L.van Genderen, Professor at the Department of Earth Observation Science, Faculty of Geo-information Science and Earth Observation(ITC),University of Twente, Netherlands, he is an Associate Editor of the international journal GEO Spatial Information Science. His research interests include urban remote sensing and new technologies for urban management. Email: [Genderen@alumni.itc.nl](mailto:Genderen@alumni.itc.nl).

# Error Analysis and Optimization of a Sky Full-Polarization Imaging Detection System

Yongtai Chen, William C. Tang, Jinkui Chu, Ran Zhang, and Song Li

## Abstract

An accurate sky polarization field map is a prerequisite for polarization navigation applications. In this article, a detector for sky full-polarization imaging detection is described, the major error-influencing factors (MEIFS) are obtained, and the error propagation is modeled and analyzed. We reveal the relationship between the error of the inversed Stokes vector and the condition number of the detector matrix, which shows that the error of the inversed Stokes vector is affected by the Stokes vector of the incident light itself and the MEIFS together, with the MEIFS playing a decisive role. With the MEIFS optimized, the impact of detector error on the inversed Stokes vector is attenuated. A control equation for system calibration is also deduced which can establish the connection between the detector matrix design and calibration process. The work in this article provides a reference for optimization and calibration of sky full-polarization imaging detectors.

## Introduction

In recent years, a novel navigation method based on the sky polarization field map (SPFM) has become a research hot spot, expected to become an important supplement to traditional navigation methods (Karman, Diah and Gebeshuber 2012; Yan *et al.* 2018). Polarization navigation is a navigation method that uses sunlight (Yan *et al.* 2009) or moonlight (Cui *et al.* 2013) as a navigation beacon to obtain compass information by detecting and calculating the SPFM (Zhao *et al.* 2007). Figure 1 shows the theoretical SPFM under Rayleigh scattering in fine weather. Polarization navigation has the properties of insusceptibility to electromagnetic interference and artificial false-signal deception, as well as good independence. Its navigation and positioning functions have wide application value in military and civilian fields (Chu *et al.* 2016). Polarization navigation is also one of the navigation methods in nature (Chu *et al.* 2007; Chu *et al.* 2008; Chu *et al.* 2009). In 1985, the first research paper was published about the polarization navigation ability of sand ants (Fent and Wehner 1985).

In 2000, Lambrinos and colleagues (2000) applied the navigation strategy of desert ants to the autonomous navigation of mobile robots and obtained satisfactory experimental results, which verified the mechanism of biological polarization-sensitive navigation. Since then, many scholars have developed

a series of polarization navigation sensors, which can be divided into two major categories: point-source polarization detectors (Chu and Zhao 2005; Cui *et al.* 2010; Liu *et al.* 2015; Wang *et al.* 2015; Chu *et al.* 2017) and imaging polarization detectors (Vedel, Breugnot and Lechocinski 2011; Chahl and Mizutani 2012; Zhang, Zhao and Li 2013; Zhang *et al.* 2014; Lu *et al.* 2015). Point-source detectors are smaller, lower in cost, and easier to integrate and modify, but the distinction they capture is imperfect, making them unsuitable for applications in complex scenes. On the other hand, imaging polarization detectors have an abundant amount of information, which is beneficial to a variety of semantic processing tasks using the polarization image. They have good potential for making polarization data mining an important technical means for obtaining an SPFM. Many scholars have used their own imaging polarization detectors to study the SPFM under various weather conditions and obtained some meaningful conclusions. However, most of the current research focuses only on the linear polarization characteristics of the SPFM; analysis of the circular polarization characteristics has been neglected (Pomozi, Horváth and Wehner 2001; Pust *et al.* 2011). There is still knowledge of the SPFM to be explored in order to make this technology more flexible and stable for application. And to enrich the database information of the SPFM for further analysis, it is beneficial to expand the detection range of the Stokes vector from linear polarization to full polarization, which requires a full-polarization imaging detection system.

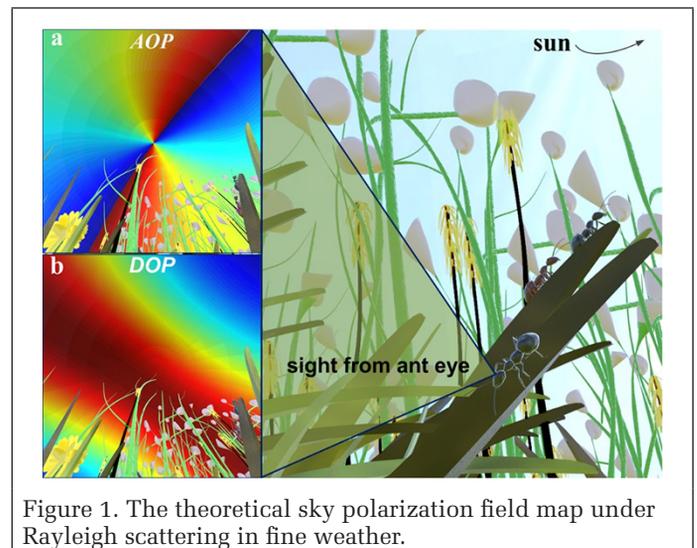


Figure 1. The theoretical sky polarization field map under Rayleigh scattering in fine weather.

Yongtai Chen is with the Key Laboratory for Micro/Nano Technology and System of Liaoning Province, Dalian University of Technology, Dalian, China; the School of Mechanical Engineering, Dalian University of Technology, Dalian, China; and the Department of Biomedical Engineering, University of California, Irvine, Irvine, CA 92697.

William C. Tang is with the Department of Biomedical Engineering, University of California, Irvine, Irvine, CA 92697.

Jinkui Chu, Ran Zhang, and Song Li are with the Key Laboratory for Micro/Nano Technology and System of Liaoning Province, Dalian University of Technology, Dalian, China; and the School of Mechanical Engineering, Dalian University of Technology, Dalian, China (chujk@dlut.edu.cn).

Photogrammetric Engineering & Remote Sensing  
Vol. 87, No. 4, April 2021, pp. 273–282.  
0099-1112/21/273–282

© 2021 American Society for Photogrammetry  
and Remote Sensing  
doi: 10.14358/PERS.87.4.273

Hence, an optimized real-time full-polarization imaging detector was developed (Chen *et al.* 2018). Compared with other detectors, it has higher integration and portability, which will help to efficiently acquire an SPFM and to provide technical support for correcting and supplementing the theoretical model of polarization navigation. This system contains several polarization modulation components, where some deviations of the theoretical nominal parameters exist owing to manufacturing error, and these will introduce systematic error in the polarization inversion processing, thus affecting the detection robustness of the SPFM. There have not been reports of comprehensive study of the systematic error formation and optimization process of this detector. Based on the needs of future work, this article devises the error model of the system and obtains some important conclusions.

## Establishment of the Detection System

### Polarization Description and Calculation

The Stokes vector with an intensity base unit is one common method to describe polarization light (Hamaoui, 2017; Lu *et al.* 2017). The Stokes vector has four parameters— $I$ ,  $Q$ ,  $U$ , and  $V$ —where  $S = [I \ Q \ U \ V]^T$ :

$$S = \begin{bmatrix} I \\ Q \\ U \\ V \end{bmatrix} = \begin{bmatrix} I_{0^\circ} + I_{90^\circ} \\ I_{0^\circ} - I_{90^\circ} \\ I_{45^\circ} - I_{135^\circ} \\ I_R - I_L \end{bmatrix}, \quad (1)$$

where  $I$  represents the total intensity of light,  $Q$  represents the intensity difference between the polarization components in the  $0^\circ$  and  $90^\circ$  directions,  $U$  represents the intensity difference between the polarization components in the  $45^\circ$  and  $135^\circ$  directions, and  $V$  represents the intensity difference between the right and left circular polarization components.

Each Stokes vector can be mapped onto the Poincaré sphere (Milione *et al.* 2011; Naidoo *et al.* 2016) as shown in Figure 2, where  $S_0$  represents the total intensity of light,  $\alpha$  is the azimuth, and  $\beta$  is the elliptic angle. Positive  $\beta$  indicates right-handed polarized light, so each point on the upper hemisphere represents right-handed polarized light, and correspondingly, each point on the lower hemispherical surface

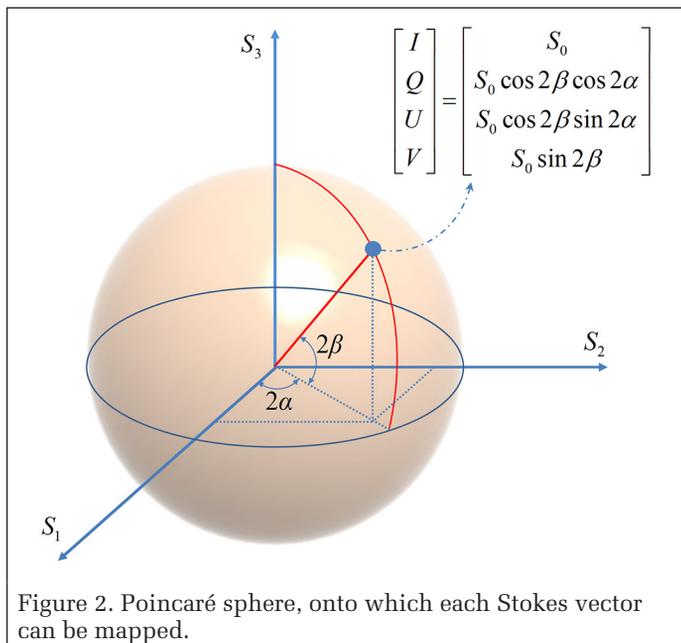


Figure 2. Poincaré sphere, onto which each Stokes vector can be mapped.

represents left-handed polarized light. Each point on the intersection of the spherical surface and the equatorial plane represents linearly polarized light, with  $\beta = 0$ . The upper and lower poles correspond to right-handed and left-handed circularly polarized light, respectively. The azimuth difference of the two polarized lights, which are symmetrical about the center of the Poincaré sphere, is  $90^\circ$ , and the absolute values of  $\beta$  are equal and with opposite signs, so a pair of orthogonally polarized lights is formed. The angle between two orthogonally linearly polarized lights is  $90^\circ$ , but on the spherical surface of the Poincaré sphere the corresponding points are located at symmetrical positions with an angle of  $\pi$  between them. The points inside the Poincaré sphere represent partially polarized lights. Thus, the points on the Poincaré sphere can represent polarized lights of any polarization state.

The four parameters of the Stokes vector can define different polarization characterization parameters (Zhang *et al.* 2015; Wu *et al.* 2018), called the SPFM when used in sky polarization detection. These are the angle of polarization (AOP), degree of polarization (DOP), degree of linear polarization (DOLP), and degree of circular polarization (DOCP), which are defined as follows:

$$\text{AOP} = \frac{1}{2} \arctan\left(\frac{U}{Q}\right) \quad (2)$$

$$\text{DOP} = \frac{\sqrt{Q^2 + U^2 + V^2}}{I} \quad (3)$$

$$\text{DOLP} = \frac{\sqrt{Q^2 + U^2}}{I} \quad (4)$$

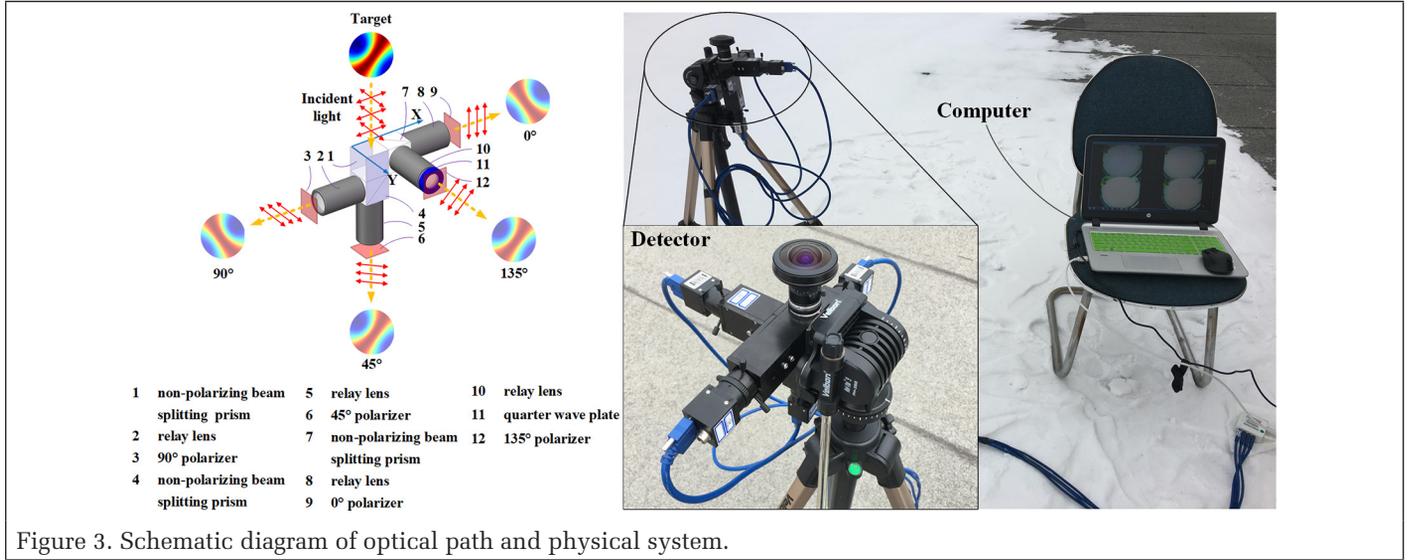
$$\text{DOCP} = \frac{\sqrt{V^2}}{I}. \quad (5)$$

### Components of the System

In order to actualize the fully rationalized Mueller matrix of the system, which can help reduce computational cost, the optical modulator components with special parameters were selected to implement the optical path, mainly including three non-polarization beam-splitting prisms (Thorlabs), four non-polarization relay lenses (Edmund Optics), one quarter-wave plate (Thorlabs), and four polarizers (Edmund Optics), as shown in Figure 3 (left). A wide-angle ( $180^\circ$ ) fish-eye lens is installed at the entrance of the optical path, and four charge-coupled device (CCD) cameras are installed at the exit to realize simultaneous image acquisition. The host computer control program, which is developed based on MFC and OpenCV, helps to implement real-time solution and analysis of polarized images. To obtain the SPFM video stream, the images from the four CCD cameras undergo a series of image processing, including solving image registration relations, registering the video stream, selecting the image color channel, solving the Stokes vector, and solving the SPFM according to Equations 2–5. The detector is smaller than a basketball, which is beneficial for rapid deployment and for platform and field switching. The physical system is shown in Figure 3 (right), and the specific parameters of the detection system are shown in Table 1.

The theoretical Mueller matrix of the polarization detection system is

$$A_0 = \begin{bmatrix} 0.125 & 0.125 & 0 & 0 \\ 0.125 & 0 & 0.125 & 0 \\ 0.125 & -0.125 & 0 & 0 \\ 0.125 & 0 & 0 & -0.125 \end{bmatrix}, \quad (6)$$



where  $i_0$ ,  $i_1$ ,  $i_2$ , and  $i_3$  represent the image intensities obtained by the four CCDs, respectively. The column vector  $I = [i_0 \ i_1 \ i_2 \ i_3]^T$  can be used to calculate the inversed Stokes vector  $S$  of the incident light:

$$S = \begin{bmatrix} I \\ Q \\ U \\ V \end{bmatrix} = A_0^{-1} \cdot I = \begin{bmatrix} 4 & 0 & 4 & 0 \\ 4 & 0 & -4 & 0 \\ -4 & 8 & -4 & 0 \\ 4 & 0 & 4 & -8 \end{bmatrix} \begin{bmatrix} i_0 \\ i_1 \\ i_2 \\ i_3 \end{bmatrix}. \quad (7)$$

## Error Analysis of the System

### Major Error-Influencing Factors (MEIFs) of the System

The introduction of systematic errors in the derivation of the system Mueller matrix is mainly caused by the manufacturing error of the components and the positional deviation in the adjusting process, thus significantly affecting the inversion accuracy of the polarization characterization parameters. There are 10 MEIFs of the system: the transmittance  $k_1$ ,  $k_2$ ,  $k_3$ , and  $k_4$  in each optical path; the polarizer mounting angles  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$ , and  $\theta_4$  in each optical path; the mounting angle  $\theta_5$  of the fast axis of the quarter-wave plate; and the phase delay  $\varphi$  of the quarter-wave plate. All the MEIFs are shown in Table 2.

The design values of all MEIFs are determined according to the ideal parameters of the optical components in the detection system. The transmittance error in each optical path mainly comes from the deviation of the beam energy in the two orthogonal directions of the non-polarization beam-splitting prism, and it is calculated by the component data sheet from the official website of Thorlabs. As there is a quarter-wave plate in the fourth optical path, as shown in Figure 3 (left), the error range of  $k_4$  is different from those of  $k_1$ ,  $k_2$ , and  $k_3$ . According to a possible misalignment of the polarizers and the quarter-wave plate during installation, an error range of  $\pm 5^\circ$  of the installation angles is produced. Based on the retardance accuracy (root-mean-square error [RMSE]) of the quarter-wave plate, the error range of the phase delay  $\varphi$  is obtained. The design values of all MEIFs help to get the fully rationalized Mueller matrix  $A_0$  of the system, but under the error impact

Table 2. The major error-influencing factors of the system.

Parameter	Factor									
	$k_1$	$k_2$	$k_3$	$k_4$	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$	$\theta_5$	$\varphi$
Design value	1/8	1/8	1/8	1/8	$0^\circ$	$45^\circ$	$90^\circ$	$135^\circ$	$0^\circ$	$\pi/2$
Error range	-0.073 to -0.0015	-0.073 to -0.0015	-0.073 to -0.0015	-0.0782 to -0.0139	$\pm 5^\circ$	$\pm \lambda/9$				

Table 1. Main performance indicators of full-sky polarization imaging detector.

Indicator	Value
Resolution (pixels)	$1024 \times 1028$
Bits/pixel	8
Field of view ( $^\circ$ )	180
Frame rate (frames/s)	3
Spectral band (nm)	450–675

of each MEIF, the actual Mueller matrix is not  $A_0$ , which can affect the accuracy of the inversed Stokes vector.

### Error Propagation

In order to obtain accurate polarization characterization parameters such as AOP, DOP, DOLP, and DQCP, it is necessary to figure out the relationship between all MEIFs and the system Mueller matrix. The detection system consists of four optical paths, each of which has its own Mueller matrix:

$$M_1 = k_1 \cdot \begin{bmatrix} 1 & \cos 2\theta_1 & \sin 2\theta_1 & 0 \\ \cos 2\theta_1 & \cos^2 2\theta_1 & \sin 2\theta_1 \cdot \cos 2\theta_1 & 0 \\ \sin 2\theta_1 & \sin 2\theta_1 \cdot \cos 2\theta_1 & \sin^2 2\theta_1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad (8)$$

$$M_2 = k_2 \cdot \begin{bmatrix} 1 & \cos 2\theta_2 & \sin 2\theta_2 & 0 \\ \cos 2\theta_2 & \cos^2 2\theta_2 & \sin 2\theta_2 \cdot \cos 2\theta_2 & 0 \\ \sin 2\theta_2 & \sin 2\theta_2 \cdot \cos 2\theta_2 & \sin^2 2\theta_2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad (9)$$

$$M_3 = k_3 \cdot \begin{bmatrix} 1 & \cos 2\theta_3 & \sin 2\theta_3 & 0 \\ \cos 2\theta_3 & \cos^2 2\theta_3 & \sin 2\theta_3 \cdot \cos 2\theta_3 & 0 \\ \sin 2\theta_3 & \sin 2\theta_3 \cdot \cos 2\theta_3 & \sin^2 2\theta_3 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad (10)$$

$$M_i = k_i \cdot \begin{bmatrix} 1 & \cos 2\theta_i & \sin 2\theta_i & 0 \\ \cos 2\theta_i & \cos^2 2\theta_i & \sin 2\theta_i \cdot \cos 2\theta_i & 0 \\ \sin 2\theta_i & \sin 2\theta_i \cdot \cos 2\theta_i & \sin^2 2\theta_i & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 - (1 - \cos \varphi) \sin^2 2\theta_i & (1 - \cos \varphi) \sin 2\theta_i \cos 2\theta_i & -\sin \varphi \sin 2\theta_i \\ 0 & (1 - \cos \varphi) \sin 2\theta_i \cos 2\theta_i & 1 - (1 - \cos \varphi) \cos^2 2\theta_i & \sin \varphi \cos 2\theta_i \\ 0 & \sin \varphi \sin 2\theta_i & -\sin \varphi \cos 2\theta_i & \cos \varphi \end{bmatrix} \quad (11)$$

$$A_{\text{error}} = \begin{bmatrix} k_1 & k_1 \cdot \cos 2\theta_1 & k_1 \cdot \sin 2\theta_1 & 0 \\ k_2 & k_2 \cdot \cos 2\theta_2 & k_2 \cdot \sin 2\theta_2 & 0 \\ k_3 & k_3 \cdot \cos 2\theta_3 & k_3 \cdot \sin 2\theta_3 & 0 \\ k_4 & m_{42} & m_{43} & m_{44} \end{bmatrix} \quad (12)$$

$$m_{42} = k_4 \cdot \cos 2\theta_4 \cdot (\sin^2 2\theta_3 \cdot (\cos \varphi - 1) + 1) - k_4 \cdot \cos 2\theta_5 \cdot \sin 2\theta_4 \cdot \sin 2\theta_5 \cdot (\cos \varphi - 1)$$

$$m_{43} = k_4 \cdot \sin 2\theta_4 \cdot (\cos^2 2\theta_3 \cdot (\cos \varphi - 1) + 1) - k_4 \cdot \cos 2\theta_4 \cdot \cos 2\theta_5 \cdot \sin 2\theta_5 \cdot (\cos \varphi - 1)$$

$$m_{44} = k_4 \cdot \cos 2\theta_5 \cdot \sin 2\theta_4 \cdot \sin \varphi - k_4 \cdot \cos 2\theta_4 \cdot \sin 2\theta_5 \cdot \sin \varphi$$

In each optical path, the Stokes vector is modulated into a light intensity signal by the Mueller matrix  $M_1$ ,  $M_2$ ,  $M_3$ , or  $M_4$  and recorded by the CCDs, respectively; then the first rows of  $M_1$ ,  $M_2$ ,  $M_3$ , and  $M_4$  are extracted to form a new  $4 \times 4$  matrix  $A_{\text{error}}$ , as shown in Equation 12. This matrix represents the actual system Mueller matrix under the impact of MEIFs. Equation 12 shows how each factor as an independent variable can affect the derivation process of the actual system Mueller matrix  $A_{\text{error}}$ .

In the real process of light transmission, the column vector  $I$  which is composed of the four image intensities is affected by  $A_{\text{error}}$ . But the theoretical Mueller matrix instead is brought into the actual inversion process, causing the error of the inversed incident Stokes vector, as follows:

$$I = A_{\text{error}} \cdot S_{\text{true}} \quad (13)$$

$$S_{\text{true}} = A_{\text{error}}^{-1} \cdot I \quad (14)$$

$$S' = A_0^{-1} \cdot I, \quad (15)$$

where  $S_{\text{true}} = [I \ Q \ U \ V]^T$  is the true Stokes vector of the incident light and  $S'$  is the inversed Stokes vector with the error.

The error transmission process of the system can be deduced from Equations 13–15:

$$\text{see equation 16 below} \quad (16)$$

where  $\Delta S = [\Delta I \ \Delta Q \ \Delta U \ \Delta V]^T$  is the error of the Stokes vector of the incident light.

Then, according to Equation 6, 7, 12, and 16, the relationship between the error in the Stokes vector of the incident light and all MEIFs can be revealed:

$$\text{see equation 17 below} \quad (17)$$

In Equation 16 and 17,  $\Delta S$  is affected by the incident Stokes vector and MEIFs together. In order to figure out the range of  $\Delta S$  that can be accepted, and to ascertain the calibration strategy of the system Mueller matrix, a simulation was conducted. The polarized light used in the simulation was full-polarization light. Through comparison analysis, we found that when the values of  $\Delta I$ ,  $\Delta Q$ ,  $\Delta U$ , and  $\Delta V$  are within  $\pm 0.005$ , the RMSE of DOP, DOLP, and DOCP are less than 0.5%, as shown in Figure 4b. As for the RMSE of AOP, 99% of it is less

$$\Delta S = S_{\text{true}} - S' = (A_{\text{error}}^{-1} - A_0^{-1}) \cdot A_{\text{error}} \cdot S_{\text{true}} = (A_{\text{error}}^{-1} \cdot A_{\text{error}} - A_0^{-1} \cdot A_{\text{error}}) \cdot S_{\text{true}} = (E - A_0^{-1} \cdot A_{\text{error}}) \cdot S_{\text{true}} = (A_0^{-1} \cdot A_0 - A_0^{-1} \cdot A_{\text{error}}) \cdot S_{\text{true}} = A_0^{-1} \cdot (A_0 - A_{\text{error}}) \cdot S_{\text{true}} = A_0^{-1} \cdot \Delta A \cdot S_{\text{true}} \quad (16)$$

$$\begin{bmatrix} \Delta I \\ \Delta Q \\ \Delta U \\ \Delta V \end{bmatrix} = \begin{bmatrix} 1 - 4 \cdot k_1 - 4 \cdot k_3 & -4 \cdot k_1 \cdot \cos 2\theta_1 - 4 \cdot k_3 \cdot \cos 2\theta_3 & -4 \cdot k_1 \cdot \sin 2\theta_1 - 4 \cdot k_3 \cdot \sin 2\theta_3 & 0 \\ -4 \cdot k_1 + 4 \cdot k_3 & 1 - 4 \cdot k_1 \cdot \cos 2\theta_1 + 4 \cdot k_3 \cdot \cos 2\theta_3 & -4 \cdot k_1 \cdot \sin 2\theta_1 + 4 \cdot k_3 \cdot \sin 2\theta_3 & 0 \\ 4 \cdot k_1 - 8 \cdot k_2 + 4 \cdot k_3 & 4 \cdot k_1 \cdot \cos 2\theta_1 - 8 \cdot k_2 \cdot \cos 2\theta_2 + 4 \cdot k_3 \cdot \cos 2\theta_3 & 1 + 4 \cdot k_1 \cdot \sin 2\theta_1 - 8 \cdot k_2 \cdot \sin 2\theta_2 + 4 \cdot k_3 \cdot \sin 2\theta_3 & 0 \\ -4 \cdot k_1 - 4 \cdot k_3 + 8 \cdot k_4 & -4 \cdot k_1 \cdot \cos 2\theta_1 - 4 \cdot k_3 \cdot \cos 2\theta_3 + 8 \cdot m_{42} & -4 \cdot k_1 \cdot \sin 2\theta_1 - 4 \cdot k_3 \cdot \sin 2\theta_3 + 8 \cdot m_{43} & 1 + 8 \cdot m_{44} \end{bmatrix} \begin{bmatrix} I \\ Q \\ U \\ V \end{bmatrix} \quad (17)$$

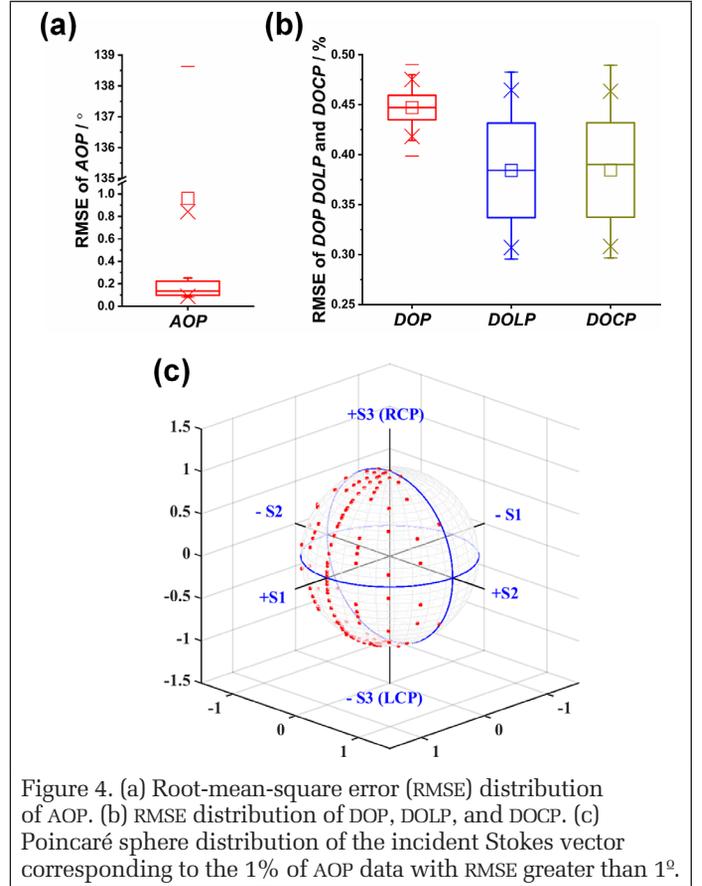


Figure 4. (a) Root-mean-square error (RMSE) distribution of AOP. (b) RMSE distribution of DOP, DOLP, and DOCP. (c) Poincaré sphere distribution of the incident Stokes vector corresponding to the 1% of AOP data with RMSE greater than  $1^\circ$ .

than  $1^\circ$ , with 75% less than  $0.25^\circ$ , as shown in Figure 4a. The incident Stokes vector corresponding to the 1% of AOP data with RMSE greater than  $1^\circ$  mainly distributes on the surface of the Poincaré sphere, where  $S_3$  is greater than 0, as shown in Figure 4c. The SPFM is one of the targets with low-frequency change (Zhang *et al.* 2015), and therefore its geometric features have a relatively high noise tolerance. The detection requirements of the SPFM can be basically satisfied by controlling the values of  $\Delta I$ ,  $\Delta Q$ ,  $\Delta U$ , and  $\Delta V$  within  $\pm 0.005$ .

In Equation 18, the 2-norm (Tyo 2002) of  $\Delta S$  represents the Euclidean distance between  $S_{\text{true}}$  and  $S'$ , and it can be used to assess the extent to which  $S'$  deviates from  $S_{\text{true}}$ :

$$D(S_{\text{true}}, S') = \|\Delta S\|_2 = \sqrt{(\Delta I)^2 + (\Delta Q)^2 + (\Delta U)^2 + (\Delta V)^2}. \quad (18)$$

The matrix condition number (Vaughn and Hoover 2008; Foster *et al.* 2015; Bruce *et al.* 2018) is the key parameter to evaluate the degree of morbidity of the system matrix;  $K(A_{\text{error}})$  is the system matrix condition number with error, it is defined as

$$K(A_{\text{error}}) = \|A_{\text{error}}\| \cdot \|A_{\text{error}}^{-1}\|. \quad (19)$$

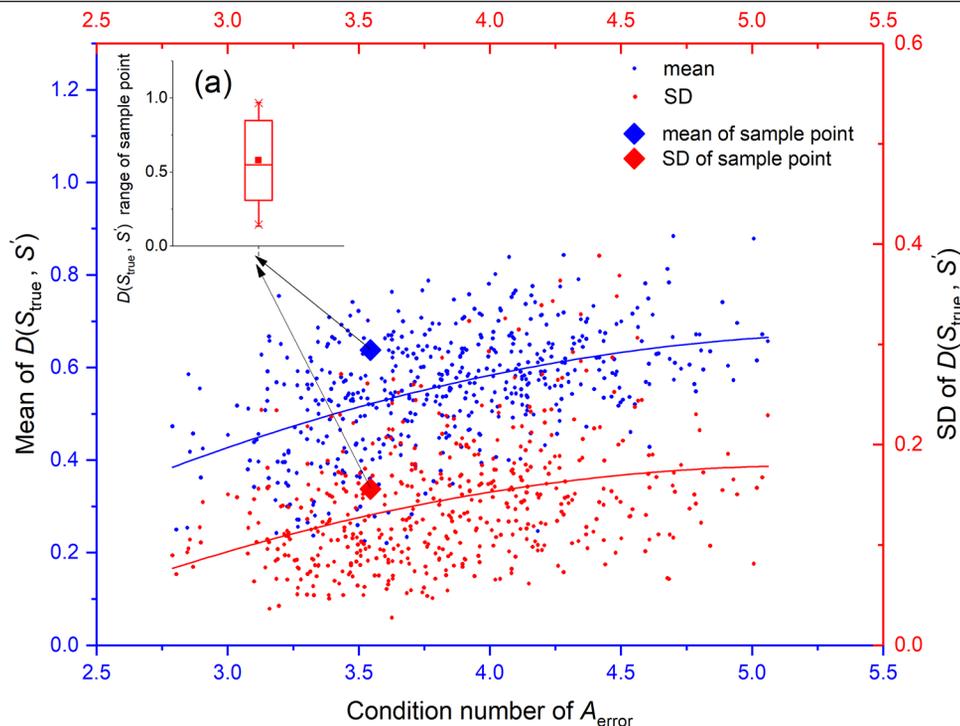


Figure 5. Influence of system matrix condition number with error and incident polarization state on detection accuracy.

There are many possible combinations of MEIFs, producing different condition numbers of  $A_{\text{error}}$ . One special situation is that different combinations of MEIFs may produce the same condition number of  $A_{\text{error}}$ . For each combination of MEIFs, the polarization state of the incident light is set using the Stokes vectors that map the entire Poincaré sphere, and the  $D(S_{\text{true}}, S')$  distribution of the inversed Stokes vector under the impact of MEIFs is computed according to Equations 16–19 and then the average value and standard deviation (SD) of  $D(S_{\text{true}}, S')$  are calculated. After analyzing the  $D(S_{\text{true}}, S')$  distribution under different MEIF combinations, we obtained the impact of the system matrix condition number with error and the incident polarization state on the detection accuracy. It can be seen in Figure 5 that the system matrix condition number with error ranges from 2.78 to 5.06, with an average of 3.84. The blue points represent the average value of  $D(S_{\text{true}}, S')$  under different matrix condition numbers, and the value range is between 0.22 and 0.88, while the red points represent the SD of  $D(S_{\text{true}}, S')$ , with values ranging from 0.02 to 0.38. The average and SD data of  $D(S_{\text{true}}, S')$  reflect that the final accuracy of the polarization detection system is affected by both MEIFs and the polarization state of the incident light.

Taking one set of data points as a sample, it can be seen from Figure 5a that  $D(S_{\text{true}}, S')$  of the sample point has a large value range, between 0.13 and 0.97, as it has a uniquely determined matrix  $A_{\text{error}}$ , and the influence on  $D(S_{\text{true}}, S')$  is mainly from the polarization state of the incident light. The respective trend lines (solid red and blue lines) are plotted according to the clustering and distribution patterns of mean and SD points, qualitatively showing that the condition number of  $A_{\text{error}}$  and the mean and SD of  $D(S_{\text{true}}, S')$  are positively correlated, which indicates that the condition number of  $A_{\text{error}}$  plays a dominant role in determining  $D(S_{\text{true}}, S')$ . Hence, it is possible to control the  $D(S_{\text{true}}, S')$  error by diminishing the condition number of the system matrix through optimizing the MEIFs. The condition number of the theoretical system matrix designed in this article is 3.2255. As shown in Figure 5, under the influence of MEIFs, about 20% of the condition numbers are smaller than the theoretical matrix condition

number, which indicates that in some combinations, MEIFs can produce a self-optimization effect on the system matrix condition number, to a maximum degree of 13.8%; however, this effect is not stable. And one condition number can correspond to different numerical distributions of  $D(S_{\text{true}}, S')$ , which indicates that the same condition number with different MEIFs has differential sensitivity to the incident polarization state. This randomness impairs the self-optimization effect of MEIFs on the system matrix, so it is necessary to avoid such randomness.

#### Optimization of System Parameters

According to the compatibility of norms, from Equation 16 we can deduce

$$D(S_{\text{true}}, S') \leq \|A_0^{-1}\|_2 \cdot \|\Delta A\|_2 \cdot \|S_{\text{true}}\|_2 \leq \sqrt{2} \cdot \|A_0^{-1}\|_2 \cdot \|\Delta A\|_2. \quad (20)$$

The maximum value of  $D(S_{\text{true}}, S')$  can be constrained by three factors, namely the 2-norms of  $A_0^{-1}$ ,  $\Delta A$ , and  $S_{\text{true}}$ , where  $\|S_{\text{true}}\|_2 \leq \sqrt{2}$  under normalized condition and  $D(S_{\text{true}}, S')$  is normalized as  $S_{\text{true}}$ . Intuitively,  $A_0$  can affect the 2-norm of  $A_0^{-1}$  and  $\Delta A$  at the same time, indicating that the degree of the randomness already mentioned is mainly caused by the theoretical Mueller matrix  $A_0$ . It is possible to control the  $D(S_{\text{true}}, S')$  variation by optimizing the MEIFs of the system matrix.

There are 10 MEIFs, and their relationship to  $D(S_{\text{true}}, S')$  is highly coupled. Considering cost control and implementation difficulty, without replacing existing system components, the optimizable parameters are mainly the mounting angles of the polarizers and quarter-wave plate  $\theta_1, \theta_2, \theta_3, \theta_4$ , and  $\theta_5$ . The flowchart of the optimization process is shown in Figure 6.

The steps of the optimization process are as follows:

1. Set  $k_1, k_2, k_3, k_4$ , and  $\varphi$  as design values, as shown in Table 2
2. Set the angular optimization ranges for  $\theta_1, \theta_2, \theta_3, \theta_4$ , and  $\theta_5$  as  $[0, \pi]$ , the step length of angular optimization for  $\theta_1, \theta_2, \theta_3, \theta_4$ , and  $\theta_5$  as  $L$ , and the number of optimizations per angle as  $\pi/L + 1$ .
3. Traverse  $\theta_1, \theta_2, \theta_3, \theta_4$ , and  $\theta_5$  in  $[0, \pi]$ :

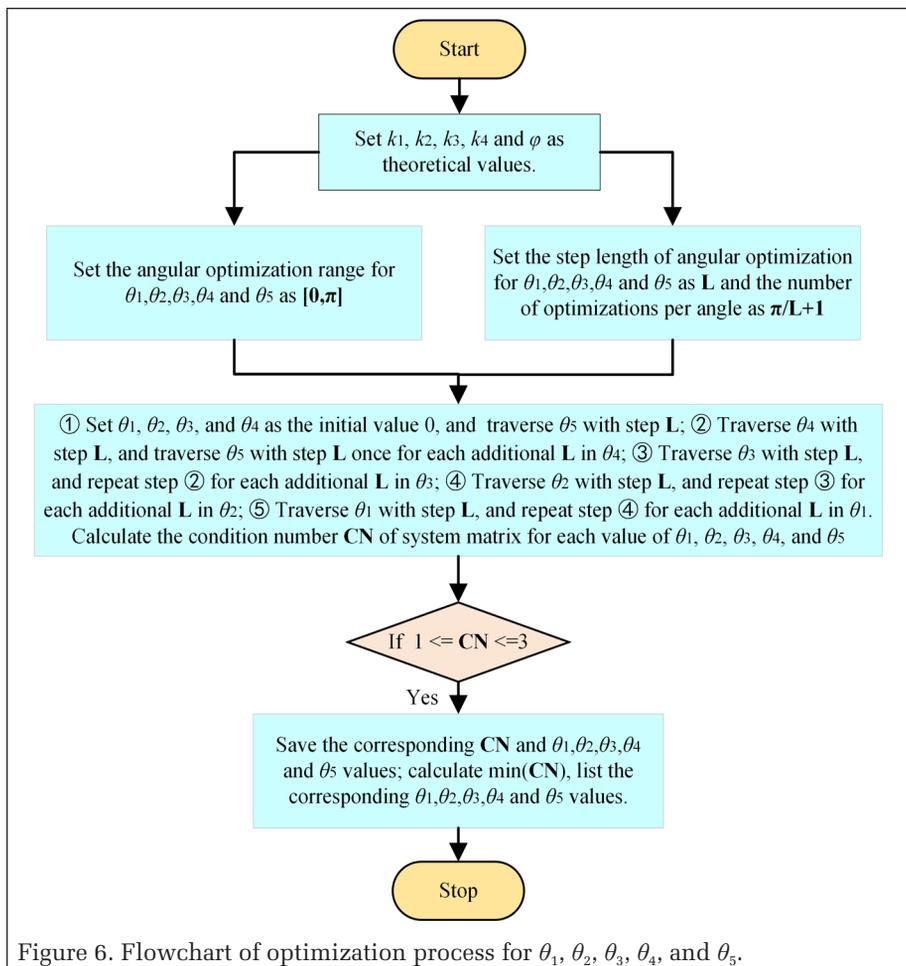


Figure 6. Flowchart of optimization process for  $\theta_1, \theta_2, \theta_3, \theta_4,$  and  $\theta_5$ .

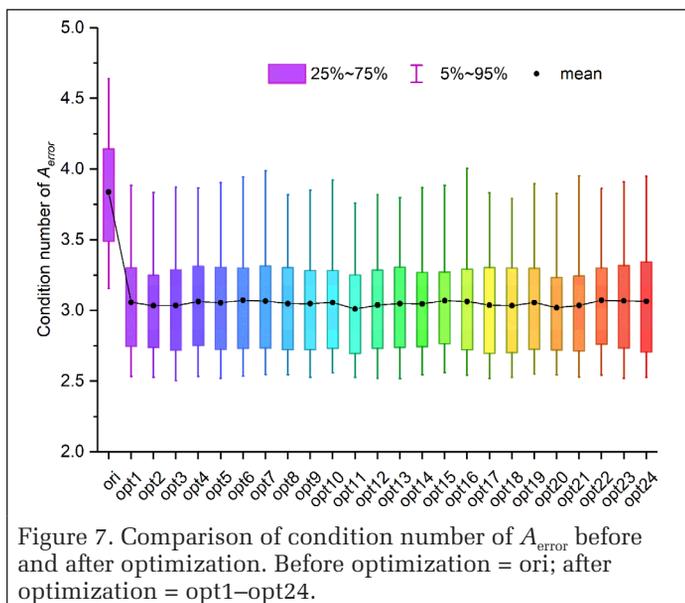


Figure 7. Comparison of condition number of  $A_{error}$  before and after optimization. Before optimization = ori; after optimization = opt1–opt24.

Table 3. 96 combinations of  $\theta_1, \theta_2, \theta_3, \theta_4,$  and  $\theta_5$  with the condition number optimized to 2.4842.

Subgroups of $(\theta_1, \theta_2, \theta_3)$	$\theta_4$	$\theta_5$
	45° or 135°	0°
	60° or 150°	15°
	75° or 165°	30°
	0° or 90°	45°
Combinations of optimized angles (0°, 60°, 120°)	15° or 105°	60°
(15°, 75°, 135°)	30° or 120°	75°
(30°, 90°, 150°)	45° or 135°	90°
(45°, 105°, 165°)	60° or 150°	105°
	75° or 165°	120°
	0° or 90°	135°
	15° or 105°	150°
	30° or 120°	165°

- e. Traverse  $\theta_1$  with step L, and repeat step (d) for each additional L in  $\theta_1$ .
  - f. Calculate the condition number CN of the system matrix for each value of  $\theta_1, \theta_2, \theta_3, \theta_4,$  and  $\theta_5$ .
4. Calculate the minimum of CN, and list the corresponding  $\theta_1, \theta_2, \theta_3, \theta_4,$  and  $\theta_5$  values.

Through this simulation, 840 sets of optimal combinations of  $\theta_1, \theta_2, \theta_3, \theta_4,$  and  $\theta_5$  values were generated. The property of the elementary transformation of a matrix indicates that after adjusting the order of the rows, a changed matrix is equivalent to the original one, and we also verified that the condition number of the matrix would not change. Equations 8–10 are formally similar, so the optimal combinations of  $\theta_1, \theta_2,$  and  $\theta_3$  can neglect their sequence;  $\theta_4$  and  $\theta_5$  are the angles of the polarizer and the quarter-wave plate in the same optical path, and they are physically sequentially related and need to be listed separately. Thus, 840 sets of the optimal combinations can be categorized into 96 sets, as shown in Table 3. In this simulation,  $\theta_5$  gets 12 optimal values, and each value can correspond to two optimal values of  $\theta_4$  and four combinations of  $\theta_1, \theta_2,$  and  $\theta_3$ . For example, when  $\theta_5 = 0^\circ$ ,  $\theta_4$  equals either  $45^\circ$  or  $135^\circ$ , and the combination of  $(\theta_1, \theta_2, \theta_3)$  can be  $(0^\circ, 60^\circ, 120^\circ), (15^\circ, 75^\circ, 135^\circ), (30^\circ, 90^\circ, 150^\circ),$  or  $(45^\circ, 105^\circ, 165^\circ)$ .

With the optimized combinations in Table 3, the condition number of the theoretical system matrix can be diminished from 3.2255 to 2.4842, a reduction of approximately 30%, and the 2-norm of  $A_0^{-1}$  is decreased from 12.0818 to 9.5808, a reduction of approximately 20.7%, which indicates that the influence of the MEIFs on  $D(S_{true}, S')$  variation can be reduced to some extent. Although the optimized combinations of  $\theta_1, \theta_2, \theta_3, \theta_4,$  and  $\theta_5$  are obtained, these ideal combinations may still be affected by the MEIFs in Table 2 during the actual assembly of the system, thus leading to biases in solving the Stokes vector as well. To this end, the effect of the MEIFs on the condition number of the actual matrix  $A_{error}$  and  $D(S_{true}, S')$  variation was simulated, with the ideal Mueller matrix  $A_0$  being the optimized one in these analyses.

Twenty-four out of 96 combinations of  $\theta_1, \theta_2, \theta_3, \theta_4,$  and  $\theta_5$  were selected for analysis, corresponding to all cases of  $\theta_4$  and  $\theta_5$  and one subgroup each of  $\theta_1, \theta_2,$  and  $\theta_3$ , respectively. The case before optimization is denoted “ori,” and opt1 through opt24 represent the cases after optimization. In Figure 7, the optimized condition number of  $A_{error}$  decreases considerably, with the average, 75% fraction, and 95% fraction all smaller

- a. Set  $\theta_1, \theta_2, \theta_3,$  and  $\theta_4$  as initial value 0 and traverse  $\theta_5$  with step L.
- b. Traverse  $\theta_4$  with step L, and traverse  $\theta_5$  with step L once for each additional L in  $\theta_4$ .
- c. Traverse  $\theta_3$  with step L, and repeat step (b) for each additional L in  $\theta_3$ .
- d. Traverse  $\theta_2$  with step L, and repeat step (c) for each additional L in  $\theta_2$ .

than before optimization, and the optimized averages are almost at the same level, indicating that the optimized ideal Mueller matrix  $A_0$  can diminish the negative effect of MEIFs on the condition number of  $A_{\text{error}}$ .

Figures 8 and 9 show the effect of MEIFs on  $D(S_{\text{true}}, S')_{\text{mean}}$  and  $D(S_{\text{true}}, S')_{\text{SD}}$  before and after optimization. In Figure 8, the optimized  $D(S_{\text{true}}, S')_{\text{mean}}$  has the same average and confidence interval, and the average is smaller than before optimization. Taking ori and opt11 as examples, the normal probability plots show that the  $D(S_{\text{true}}, S')_{\text{mean}}$  for opt11 as a whole indicates a relatively smaller value distribution than for ori. In Figure 9, the average values of almost all optimized  $D(S_{\text{true}}, S')_{\text{SD}}$  are smaller than before optimization, and the normal probability plots also show that the  $D(S_{\text{true}}, S')_{\text{SD}}$  for opt11 has a relatively smaller numerical interval than for ori. Figures 8 and 9 both indicate that the optimized MEIFs have an attenuated effect on  $D(S_{\text{true}}, S')$  to some extent.

### Relationship Between System Design and Calibration

When the values of  $\Delta I$ ,  $\Delta Q$ ,  $\Delta U$ , and  $\Delta V$  are within  $\pm 0.005$ ,  $D(S_{\text{true}}, S')_{\text{max}} = 0.01$ . Using  $D(S_{\text{true}}, S')$  to assess the extent to which  $S'$  deviates from  $S_{\text{true}}$ , the uncertainty of the precision control of the incident Stokes vector is replaced by a certain value  $D(S_{\text{true}}, S')_{\text{max}}$ . The calibration accuracy control formula can be deduced from Equation 20 as

$$\|\Delta A\|_2 \leq \frac{D(S_{\text{true}}, S')_{\text{max}}}{\sqrt{2} \cdot \|A_0^{-1}\|_2} = \frac{0.01}{\sqrt{2} \cdot \|A_0^{-1}\|_2}. \quad (21)$$

Equation 21 contains three variables:  $\|A_0^{-1}\|_2$  reflects the design accuracy of the polarization detection system,  $\|\Delta A\|_2$  reflects the control accuracy of the calibration system, and  $D(S_{\text{true}}, S')_{\text{max}}$  reflects the maximum value of  $D(S_{\text{true}}, S')$ . The equation links the two mutually independent design processes of the polarization detection system and the calibration system. Designing and implementing the calibration system according to Equation 21 can help to examine the required accuracy of the polarization detection system, and the value of  $D(S_{\text{true}}, S')_{\text{max}}$  can also be set according to the inversed Stokes vector accuracy.

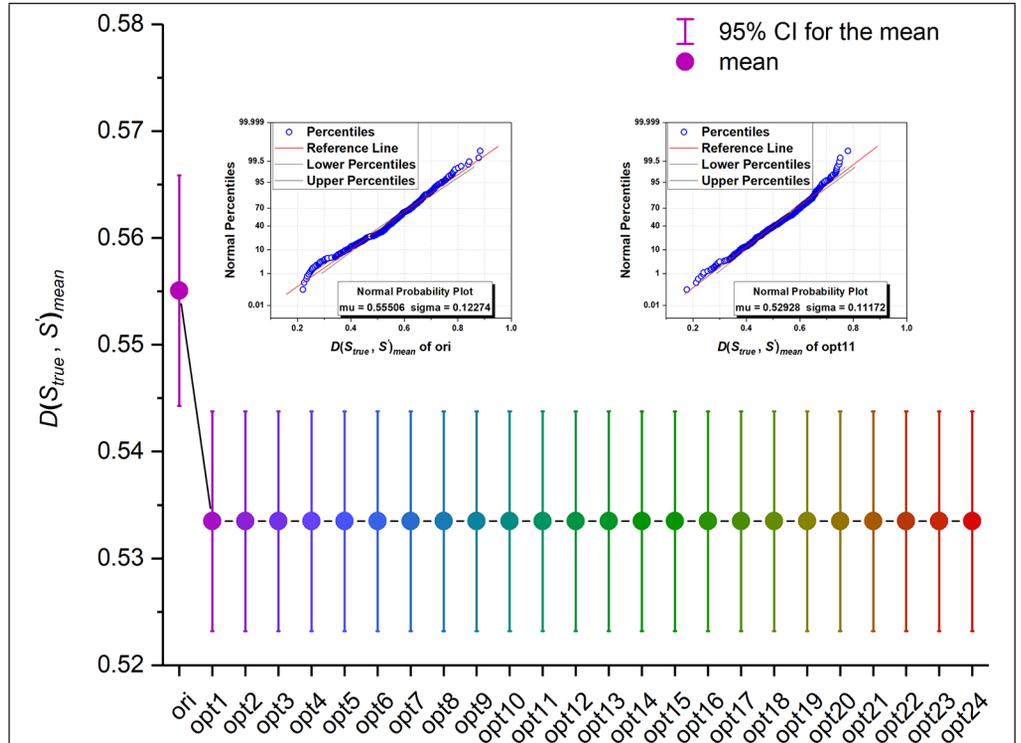


Figure 8. Comparison of  $D(S_{\text{true}}, S')_{\text{mean}}$  before and after optimization.  $D(S_{\text{true}}, S')_{\text{mean}}$  is the average of  $D(S_{\text{true}}, S')$ . Before optimization = ori; after optimization = opt1–opt24.

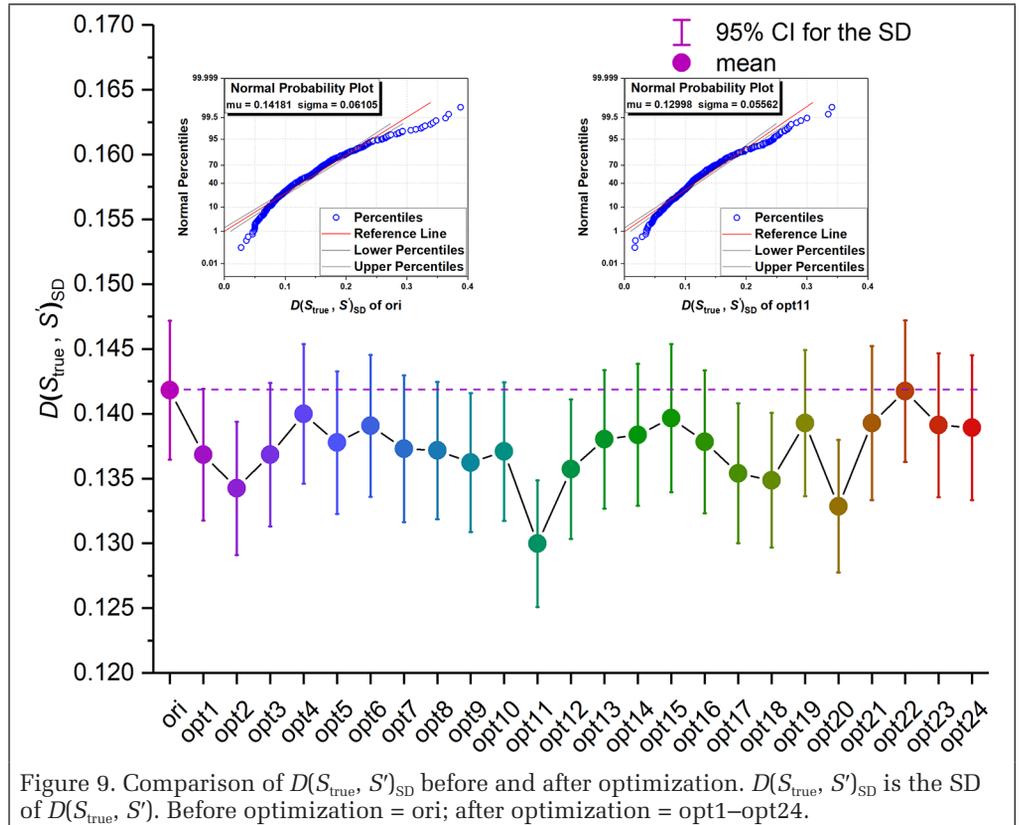


Figure 9. Comparison of  $D(S_{\text{true}}, S')_{\text{SD}}$  before and after optimization.  $D(S_{\text{true}}, S')_{\text{SD}}$  is the SD of  $D(S_{\text{true}}, S')$ . Before optimization = ori; after optimization = opt1–opt24.

## Calibration Analysis and Experiments

### Calibration Analysis

Equation 16 and Figure 5 show that the accuracy of the inversed Stokes vector is affected by the incident polarization state itself and the MEIFs together. Equation 21 is deduced and the

calibration accuracy control formula is obtained. In Equation 21,  $\|\Delta A\|_2$  reflects the control accuracy of the calibration system. However,  $\|\Delta A\|_2$  can be further divided into two parts:

$$\Delta A = \Delta A_m + \Delta A_c \quad (22)$$

$$\|\Delta A\|_2 = \|\Delta A_m + \Delta A_c\|_2 \leq \|\Delta A_m\|_2 + \|\Delta A_c\|_2, \quad (23)$$

where  $\Delta A_m$  represents the error introduced by the optics components in the detection system and  $\Delta A_c$  represents the error introduced by the calibration system. According to Equations 21–23), with a certain  $D(S_{true}, S')_{max}$ , the smaller the  $\|A_0^{-1}\|_2$ , the lower the limit of  $\|\Delta A\|_2$ , which means that the error tolerance of the detection system and the calibration system is higher.

Technically, one way to reduce  $\Delta A_m$  is to precisely adjust the optics components in the detection system according to the theoretical parameters, but this is difficult. The value of  $\Delta A_c$  is mainly generated from the light source section of the calibration system. For a full-polarization calibration system, the polarized light generator mainly includes an integrating sphere, a polarizer, and a wave plate. A variety of polarized incident light can be generated by changing the polarizing angle  $\zeta$  of the polarizer and the fast axis angle  $\eta$  of the wave plate. However,  $\zeta$  and  $\eta$  will dynamically change in operation, and even a small  $\Delta\zeta$  or  $\Delta\eta$  can change the polarization state of the incident light and introduce calibration error in the fitting solution of the Mueller matrix. In order to eliminate dynamic error of  $\zeta$  and  $\eta$ , high-precision motorized turntables are required to rotate the polarizer and wave plate. Also,  $\Delta\zeta$ ,  $\Delta\eta$ , and the initial values of  $\zeta_0$  and  $\eta_0$  should be treated as parameters to be calibrated, which can be fitted together with the Mueller matrix of the detection system to eliminate  $\Delta A_c$  and  $\Delta A_m$ . Based on these considerations, the method used in the calibration was the nonlinear least-squares fit (NLSF), which is a tool for solving multi-parameter nonlinear problems.

## Experiments

The nonlinear least-squares fit method was used to calibrate the system. The calibration process includes an integrating sphere light source, a modulation unit of the incident light polarization, and the full-polarization imaging detection system. The modulation unit consists of a 532-nm filter, a high-precision motorized turntable, a polarizer, a precision angle sensor, and a quarter-wave plate, in sequence. The polarizer is mounted on the turntable, and the quarter-wave plate is on the angle sensor. The main calibration steps are as follows:

1. There are 10 MEIFs in the detection system, and the purpose of the calibration is to get the accurate detector system matrix  $M$ , so the 16 elements of  $M$  can represent the MEIFs in the calculation. Then 20 unknown parameters of the calibration process are determined: the 16 elements of the detector system matrix  $M$  and four other parameters of the modulation unit—the transmittance  $\tau$ , the initial angles  $\zeta_0$  and  $\eta_0$  of the

polarizer and quarter-wave plate, and the phase delay  $\Delta$  of the quarter-wave plate.

2. Rotate the quarter-wave plate in the modulation unit from  $0^\circ$  to  $360^\circ$  at  $45^\circ$  intervals. With each angle of the quarter-wave plate, rotate the polarizer from  $0^\circ$  to  $360^\circ$  at  $45^\circ$  intervals to create polarized incident light, and record all 64 sets of data with the detector.
3. Applying the nonlinear least-squares fitting method, fit the above 20 parameters.
4. According to the fitted initial angle  $\zeta_0$  of the polarizer, use the high-precision motorized turntable to adjust the polarizer direction to  $0^\circ$ . Remove the quarter-wave plate from the modulation unit, vary the polarizer direction from  $0^\circ$  to  $170^\circ$ , and record every  $10^\circ$  by the detection system, for a total of 18 sets.
5. Add the quarter-wave plate and angle sensor behind the polarizer and adjust the direction of the fast axis of the wave plate to  $0^\circ$  according to the fitted value  $\eta_0$ . Adjust the polarizer direction to  $-45^\circ$  to create the right-hand circularly polarized light, and record it by the detection system.
6. Repeat (5) for a total of seven times.
7. Calculate AOP, DOLP, and DOCP of the linearly and circularly polarized incident light using the system matrix before and after calibration. Compare the errors with the created known polarized light. Results are shown in Figures 10–12.

Figures 10 and 11 show the AOP and DOLP of the linearly polarized incident light calculated by the detection system before and after the NLSF calibration. The detection errors of AOP and DOLP are less than  $0.2^\circ$  and 2.5%, respectively, after the calibration. Figure 12 shows the DOCP of right-handed circularly polarized incident light calculated by the detection system before and after the NLSF calibration. The detection error of DOCP is less than 3.6% after the calibration. The errors in AOP, DOLP, and DOCP are all reduced by approximately

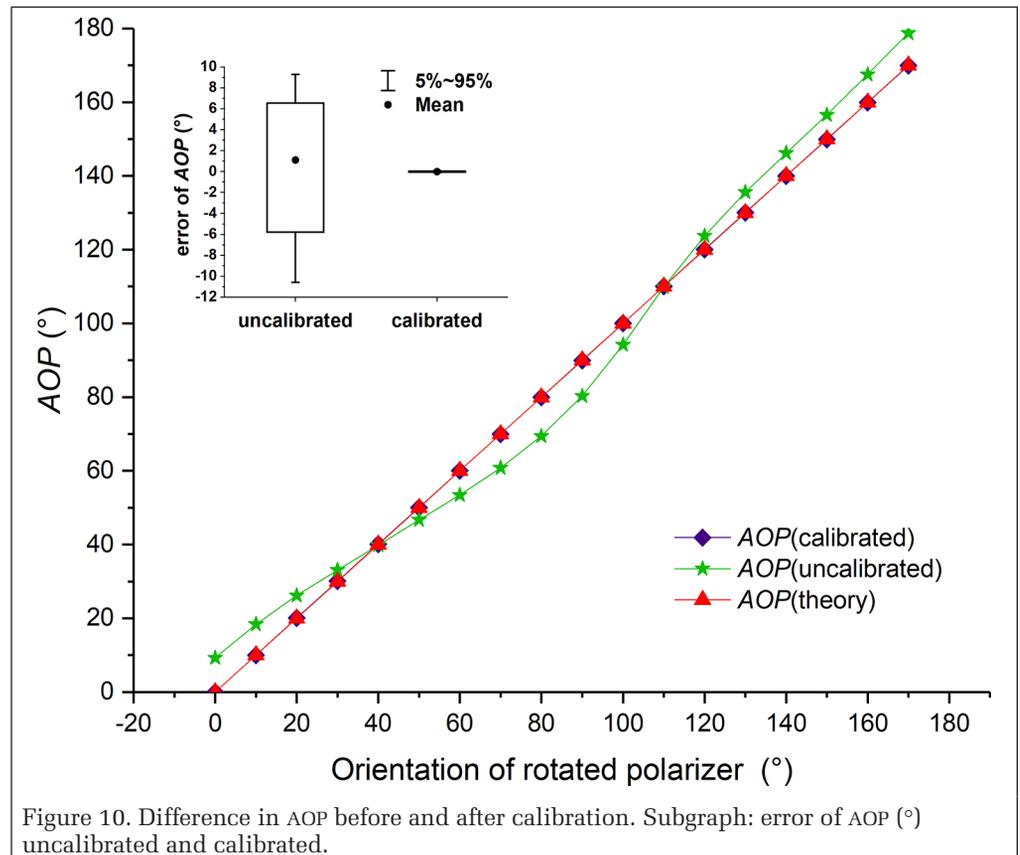


Figure 10. Difference in AOP before and after calibration. Subgraph: error of AOP ( $^\circ$ ) uncalibrated and calibrated.

50% with the NLSF method, compared with the calibration method without  $\zeta$  and  $\eta$  as fitting parameters.

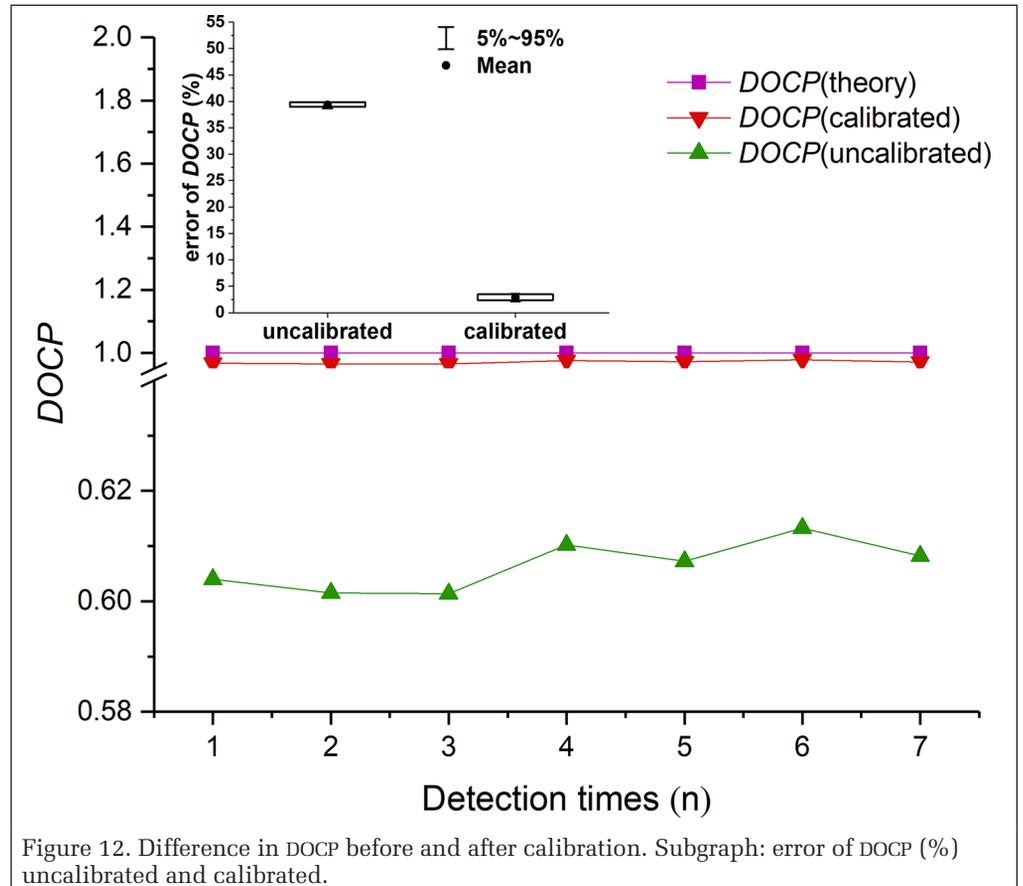
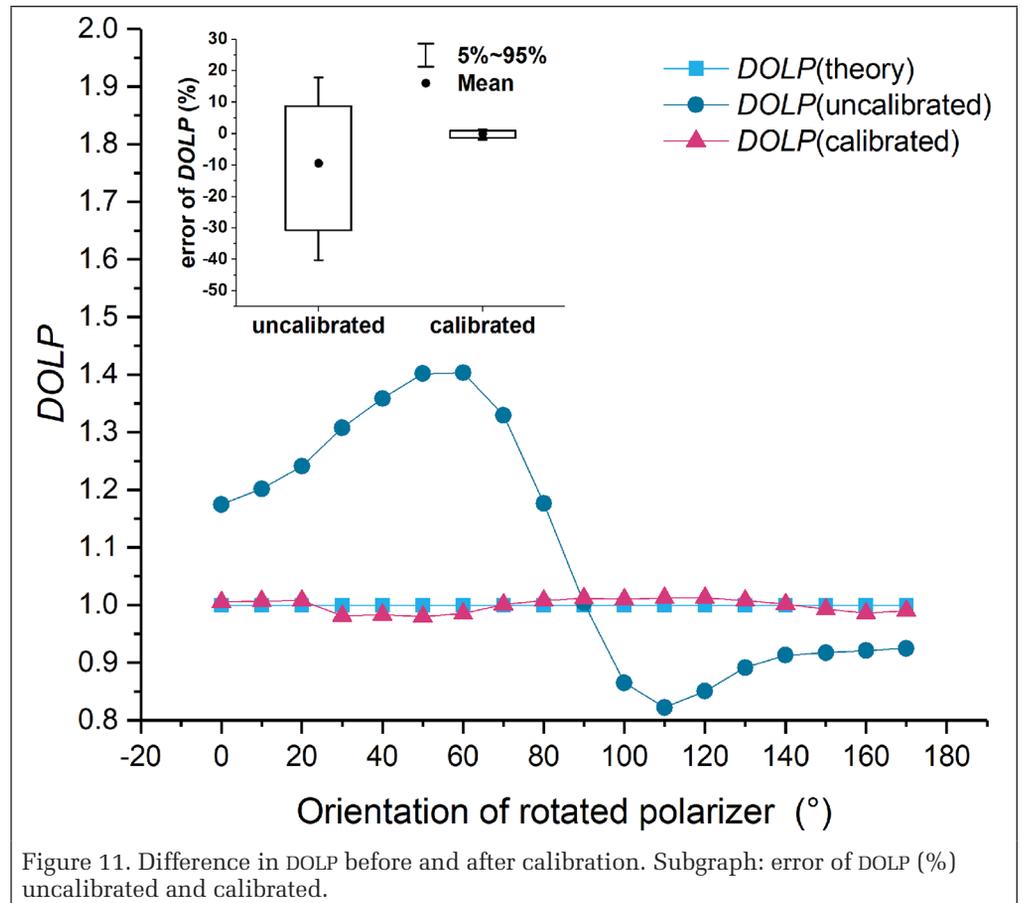
### Conclusions and Future Work

In this article, the MEIFs of the full-polarization detection system were analyzed, the error propagation process was derived, and the detection system was calibrated. In summary, the error propagation process indicates that the error of the inversed Stokes vector is affected both by MEIFs and by the polarization state of the incident light. By applying the 2-norm of the Stokes vector and the condition number of the system Mueller matrix as the evaluation criteria of detection accuracy and MEIF impact, respectively, we found that MEIFs dominate the formation of detection error, and their relationship is positively correlated, so the detection error can be controlled by attenuating the condition number of the system matrix while designing the detection system.

After the optimization of  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$ ,  $\theta_4$ , and  $\theta_5$ , 96 sets of angle combinations were obtained with smaller matrix condition numbers, and the analyses indicate that the optimal combinations can diminish the negative effect of MEIFs and the polarization state of the incident light on the detection accuracy. Under the acceptable error range of the inversed Stokes vector, the control equation of system calibration was ascertained, which can help to examine the design and calibration of the detection system. The work in this article provides a reference for the optimization and calibration of full-polarization imaging detectors.

### Acknowledgments

This work was supported by the National Natural Science Foundation of China (51675076, 51505062), the Science Fund for Creative Research Groups of the National Natural Science Foundation of China (51621064), and the Pre-Research Foundation of China (61405180102).



## References

- Bruce, N. C., J. M. López-Téllez, O Rodríguez-Núñez and O. G. Rodríguez-Herrera. 2018. Permitted experimental errors for optimized variable-retarder Mueller-matrix polarimeters. *Optics Express* 26 (11):13693–13704.
- Chahl, J. and A. Mizutani. 2012. Biomimetic attitude and orientation sensors. *IEEE Sensors Journal* 12 (2):289–297.
- Chen, Y.-T., R. Zhang, W. Lin and J.-K. Chu. 2018. Design and construction of real-time all-polarization imaging detector for skylight. *Optics and Precision Engineering* 26 (4):816–824.
- Chu, J., H. Wang, W. Chen and R. Li. 2009. Application of a novel polarization sensor to mobile robot navigation. Pages 3763–3768 in *2009 International Conference on Mechatronics and Automation*, held in Changchun, China, 9–12 August 2009. Edited by J. Editor. City, St.: IEEE.
- Chu, J.-K., H.-X. Zhang, Y.-L. Wang and C. Shi. 2017. Design and construction of autonomous real-time position prototype based on multi-polarized skylight. *Optics and Precision Engineering* 25 (2):312–318.
- Chu, J., R. Zhang, Z. Wang and Y. Wang. 2016. Progress on bio-inspired polarized skylight navigation sensor. *Chinese Science Bulletin* 61 (23):2568–2577.
- Chu, J.-K. and K.-C. Zhao. 2005. Study of angle measurement optoelectronic model on emulating polarization-sensitive compound eye of insect. *Micronanoelectronic Technology* 42 (12):541–545.
- Chu, J., K. Zhao, T. Wang and Q. Zhang. 2007. Research on a novel polarization sensor for navigation. Pages 241–246 in *2007 International Conference on Information Acquisition*, held in Seogwipo-si, South Korea, 8–11 July 2007. Edited by J. Editor. City, St.: IEEE.
- Chu, J., K. Zhao, Q. Zhang and T. Wang. 2008. Construction and performance test of a novel polarization sensor for navigation. *Sensors and Actuators A: Physical* 148 (1):75–82.
- Cui, Y., J. K. Chu, N. N. Cao and K. C. Zhao. 2010. Study on measuring system for characteristics and distribution of skylight polarization. *Key Engineering Materials* 437:369–373.
- Cui, Y., Q.-S. Gao, J.-K. Chu and C. Chen. 2013. Influence of sunlight and moonlight on polarization patterns during twilight. *Optics and Precision Engineering* 21 (1):34–39.
- Fent, K. and R. Wehner. 1985. Ocelli: a celestial compass in the desert ant *Cataglyphis*. *Science* 228 (4696):192–194.
- Foster, G., A. Karastergiou, R. Paulin, T. D. Carozzi, S. Johnston and W. van Straten. 2015. Intrinsic instrumental polarization and high-precision pulsar timing. *Monthly Notices of the Royal Astronomical Society* 453 (2):1489–1502.
- Hamaoui, M. 2017. Polarized skylight navigation. *Applied Optics* 56 (3):B37–B46.
- Karman, S. B., S.Z.M. Diah and I. C. Gebeshuber. 2012. Bio-inspired polarized skylight-based navigation sensors: a review. *Sensors* 12 (11):14232–14261.
- Lambrinos, D., R. Möller, T. Labhart, R. Pfeifer and R. Wehner. 2000. A mobile robot employing insect strategies for navigation. *Robotics and Autonomous Systems* 30 (1–2):39–64.
- Liu, Z., R. Zhang, Z. Wang, L. Guan, B. Li and J. Chu. 2015. Integrated polarization-dependent sensor for autonomous navigation. *Journal of Micro/Nanolithography, MEMS, and MOEMS* 14 (1):015001.
- Lu, H., K. Zhao, Z. You and K. Huang. 2015. Angle algorithm based on Hough transform for imaging polarization navigation sensor. *Optics Express* 23 (6):7248–7262.
- Lu, H., K. Zhao, Z. You and K. Huang. 2017. Real-time polarization imaging algorithm for camera-based polarization navigation sensors. *Applied Optics* 56 (11):3199–3205.
- Milione, G., H. I. Sztul, D. A. Nolan and R. R. Alfano. 2011. Higher-order Poincaré sphere, Stokes parameters, and the angular momentum of light. *Physical Review Letters* 107 (5):053601.
- Pomozzi, I., G. Horváth and R. Wehner. 2001. How the clear-sky angle of polarization pattern continues underneath clouds: full-sky measurements and implications for animal orientation. *Journal of Experimental Biology* 204 (17):2933–2942.
- Pust, N. J., A. R. Dahlberg, M. J. Thomas and J. A. Shaw. 2011. Comparison of full-sky polarization and radiance observations to radiative transfer simulations which employ AERONET products. *Optics Express* 19 (19):18602–18613.
- Tyo, J. S. 2002. Design of optimal polarimeters: maximization of signal-to-noise ratio and minimization of systematic error. *Applied Optics* 41 (4):619–630.
- Vaughn, I. J. and B. G. Hoover. 2008. Noise reduction in a laser polarimeter based on discrete waveFigure rotations. *Optics Express* 16 (3):2091–2108.
- Vedel, M., S. Breugnot and N. Lechocinski. 2011. Full Stokes polarization imaging camera. *Proceedings of SPIE* 8160:81600X.
- Wang, Y., J. Chu, R. Zhang, L. Wang and Z. Wang. 2015. A novel autonomous real-time position method based on polarized light and geomagnetic field. *Scientific Reports* 5:9725.
- Wu, T., Q. Tong, P. Pellikka, C. Zhang and L. Yan. 2018. Neutral point separation method for polarized effect between land objects and atmosphere in polarization remote sensing. *Journal of Remote Sensing* 22 (6):980–988.
- Yan, L., X. Gu, J. Chu, Z. You, S. Liu, M. Hugh and V. Chandrasekar. 2018. Optical polarized effects for high-resolution quantitative remote sensing and new polarization remote sensing fields. *Journal of Remote Sensing* 22 (6):901–916.
- Yan, L., G. X. Guan, J. B. Chen, T. X. Wu and X. Shao. 2009. The bionic orientation mechanism in the skylight polarization pattern. *Acta Scientiarum Naturalium Universitatis Pekinensis* 45 (4):616–620.
- Zhang, W., Y. Cao, X. Zhang and Z. Liu. 2015. Sky light polarization detection with linear polarizer triplet in light field camera inspired by insect vision. *Applied Optics* 54 (30):8962–8970.
- Zhang, Y., H. Zhao and N. Li. 2013. Polarization calibration with large apertures in full field of view for a full Stokes imaging polarimeter based on liquid-crystal variable retarders. *Applied Optics* 52 (6):1284–1292.
- Zhang, Y., H. Zhao, P. Song, S. Shi, W. Xu and X. Liang. 2014. Ground-based full-sky imaging polarimeter based on liquid crystal variable retarders. *Optics Express* 22 (7):8749–8764.
- Zhao, K., J. K. Chu, H. Y. Yao, Q. Zhang and T. S. Wang. 2007. Simulation and prediction for Rayleigh skylight polarization distribution. *Journal of Sichuan University: Engineering Science Edition* 39 (S1):287–291.

# Parsing of Urban Facades from 3D Point Clouds Based on a Novel Multi-View Domain

Wei Wang, Yuan Xu, Yingchao Ren, and Gang Wang

## Abstract

Recently, performance improvement in facade parsing from 3D point clouds has been brought about by designing more complex network structures, which cost huge computing resources and do not take full advantage of prior knowledge of facade structure. Instead, from the perspective of data distribution, we construct a new hierarchical mesh multi-view data domain based on the characteristics of facade objects to achieve fusion of deep-learning models and prior knowledge, thereby significantly improving segmentation accuracy. We comprehensively evaluate the current mainstream method on the RueMonge 2014 data set and demonstrate the superiority of our method. The mean intersection-over-union index on the facade-parsing task reached 76.41%, which is 2.75% higher than the current best result. In addition, through comparative experiments, the reasons for the performance improvement of the proposed method are further analyzed.

## Introduction

Facade parsing is an important part of 3D city modeling. It directly determines efficient constructs of detailed 3D urban models with semantic details, such as building models at level of detail 3 in CityGML format (Gröger and Plümer 2012; Biljecki *et al.* 2014; Biljecki *et al.* 2016). Facade parsing not only represents the graphic appearance of a building model but also contains the attributes and semantic information of the building (windows, walls, roofs, etc.). There are important applications in the field of smart cities and urban reconstruction (M. Zhang *et al.* 2013; Wendel *et al.* 2017; Shao *et al.* 2016; Saretta *et al.* 2020). In recent years, with the improvement of photogrammetry and visual-reconstruction technology, we can easily obtain sparse point-cloud data with color attributes based on structure from motion (SfM) at a low cost.

In SfM data, the 3D point cloud with depth information is a good representation of the hierarchical layout features of the facade, and the 2D texture information can express the architectural design style very well. However, it is still challenging to parse the facade structure from the SfM point cloud with high precision and automation. The difficulty of the problem is mainly due to two factors: the point-cloud data reconstructed based on SfM are sparse, and the structure disordered; and facade design styles are very different, with complicated layouts. Among the current mainstream methods, the classic methods based on rule constraints and geometric features (Zheng *et al.* 2010; Y. Li *et al.* 2011; Wan and Sharf 2012;

Martinović *et al.* 2015; Z. Li *et al.* 2017; Gadde *et al.* 2018) are sensitive to point-cloud density and architectural style types but are not robust, whereas methods based on deep learning do not effectively use the facade structure and other prior knowledge in the model design (Charles *et al.* 2017; Qi *et al.* 2017; Su *et al.* 2018; Liu *et al.* 2019).

In this article, we combine prior knowledge of facades with semantic segmentation methods of multi-view projection to propose a new method based on a hierarchical mesh multi-view (HMMV) domain, which can automatically achieve semantic segmentation of the facade from the SfM point cloud. The main contributions of this article are as follows:

1. We develop a hierarchical mesh (H-Mesh), a new 3D facade mesh model that is capable of extracting prior structure knowledge of the facade. Due to the hierarchical characteristics of various elements in the distribution of facade depth-attribute values, based on this prior knowledge this article performs fuzzy clustering on the original mesh to generate the H-Mesh, which can significantly characterize the edge information of the facade elements.
2. We create the HMMV domain, a new 2D facade data domain, through 2D projection and sampling of the H-Mesh. The HMMV domain also integrates facade texture information and edge information to enable a deep segmentation network to learn more discriminative features and improve the accuracy of semantic segmentation. To make the HMMV domain express the 3D H-Mesh as accurately as possible, we design a multi-angle, multi-scale sampling strategy.
3. Based on the previous steps, we propose a new facade point-cloud parsing framework. Without the use of additional data sets or data enhancements, the proposed method achieves the best results in all performance indicators on the RueMonge 2014 data set 3D facade-parsing task: overall pixel accuracy, average class accuracy, and mean intersection over union.

The organization of this article is as follows: In the next section, we introduce the related methods of facade parsing. After that, we describe the method of creating the H-Mesh, the process of generating the HMMV domain, semantic segmentation, and the details of back projection. Then we introduce the experimental data set, implementation details, and qualitative and quantitative analysis of the experimental results. Finally, we summarize the article.

## Related Work

The focus of this article is on facade parsing with a point cloud. The goal is to assign semantic labels (windows, balconies, walls, etc.) to each point. This work is closely related to the field of 3D point-cloud segmentation. We divide the related work into two categories of solution techniques:

Photogrammetric Engineering & Remote Sensing  
Vol. 87, No. 4, April 2021, pp. 283–293.  
0099-1112/21/283–293

© 2021 American Society for Photogrammetry  
and Remote Sensing  
doi: 10.14358/PERS.87.4.283

Wei Wang and Yingchao Ren are with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China; and the College of Resources and Environment, University of Chinese Academy of Sciences, Beijing, China (renyc@radi.ac.cn).

Yuan Xu is with the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, China.

Gang Wang is with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China.

knowledge-driven and data-driven. Knowledge-driven segmentation methods are based on features and prior knowledge, whereas data-driven methods are based on deep learning.

### Methods Based on Features and Prior Knowledge

The use of 3D features and prior knowledge of facades in a point cloud is the key to 3D facade semantic segmentation tasks. The feature descriptor forms a vector or matrix from a single point cloud and its local field through a specific calculation, then inputs this to the classifier to complete the point-cloud classification. Anguelov *et al.* (2005), Ladický *et al.* (2012), and Munoz *et al.* (2012) have combined a variety of 3D local and domain features and 2D texture features to achieve better segmentation results. L. Zhang *et al.* (2018) used the ReLU-NN classifier to accomplish the semantic annotation of a point cloud by constructing a feature-descriptor submatrix.

Since the layout of an architectural facade is symmetrical and hierarchical, based on this feature repetitive units are extracted and enhanced by merging them, which can improve semantic segmentation performance (Zheng *et al.* 2010; Y. Li *et al.* 2011; Z. Li *et al.* 2017). Shape grammar is a type of regularized prior knowledge. According to the structure-aware characteristics of facades, corresponding grammatical rules are predefined to guide subsequent parsing tasks (Teboul *et al.* 2011; Riemenschneider *et al.* 2012; Wan and Sharf 2012; Martinovic and Van Gool 2013; Kozi ski *et al.* 2015). Although we can obtain better parsing results based on these grammar rules, the difficulty of designing the grammar is high, which severely restricts the application scope of this method. Gadde *et al.* (2018) applied context prior knowledge to the classification of facades in 2D images and 3D point cloud by using auto-context features to train a series of boosted decision trees to achieve the semantic segmentation. Martinović *et al.* (2015) implemented a three-stage method for parsing facade point clouds; however, over-segmentation of facade instances will lead to poor facade-parsing performance.

### Methods Based on Deep Learning

There are many methods for the semantic segmentation of point clouds based on deep learning, which can be roughly divided into four categories. Among them, projection-based methods (Su *et al.* 2015; Lawin *et al.* 2017; Boulch *et al.* 2018) and discretization-based methods (Jing and You 2016; Riegler *et al.* 2017; Su *et al.* 2018; Meng *et al.* 2019) both transform irregular point-cloud structures into regular expressions in the intermediate stage and then project the intermediate-stage segmentation results back to the original point cloud. In the semantic feature-extraction stage, some special feature-learning methods have been proposed to improve performance (Lin *et al.* 2017; Wang *et al.* 2018; Shao *et al.* 2019; Shao *et al.* 2020). Point-based methods (Charles *et al.* 2017; Qi *et al.* 2017) are directly processed on the irregular point cloud. The main purpose of hybrid methods (Shao *et al.* 2017; Dai and Nießner 2018; Chiang *et al.* 2019; Jaritz *et al.* 2019; Liu *et al.* 2019) is to aggregate multi-modal features to improve the prediction ability of the model. The representative work is summarized in the following.

#### Projection-Based Methods

MVCNN (Multi-View Convolutional Neural Networks; Su *et al.* 2015) is a pioneering work, which first proposed multi-view features and applied them to 3D shape recognition. Similarly, SnapNet (Boulch *et al.* 2018) randomly acquires a large number of views in the scene and generates RGB and depth snapshots, then uses a 2D segmentation network to perform pixel-wise labeling on these snapshots. Different from SnapNet random sampling, our proposed method samples the building facade at multiple scales and angles. In addition, SnapNet directly takes a snapshot of the original RGB mesh, which is suitable for large-scale scenes but not for objects with

fine-grained structures, like facades. The generated views are weak in expressing the local details of the facade. In contrast, the multi-views generated in this article based on the H-Mesh model integrate the edge information of the facade, which can better represent the prior knowledge of facade structures.

#### Discretization-Based Methods

Su *et al.* (2018) used a sparse bilateral convolutional layer to construct sparse lattice networks (SplatNet). The method first interpolates the original point cloud to the sparse lattice of a tetrahedron, and then the sparse lattice is transformed by the bilateral convolutional layer. The occupied part is convolved, and then the filtered output value is interpolated back to the original point cloud. Riegler *et al.* (2017) designed OctNet, a 3D convolutional neural network based on an octree and a grid structure, to achieve a deeper network without affecting the resolution. Although OctNet can achieve high-resolution representation of sparse point clouds through an unbalanced octree structure, the 3D octree grid will not only cause information loss when dividing the space but will also inevitably destroy the fine-grained structure of the facade. Our proposed H-Mesh model can retain the local detail features of the facade by losing a small number of fuzzy points.

#### Point-Based Methods

PointNet (Charles *et al.* 2017) is the first deep network to manipulate point clouds directly. It learns point-wise features through a parameter-sharing multi-layer perceptron network structure and uses symmetric functions to learn global features; however, it does not extract local features. To solve this problem, PointNet++ was proposed (Qi *et al.* 2017), which aggregates the point cloud layer by layer, gradually increases the local receptive field, and adopts multi-scale and multi-resolution aggregation methods to reduce the influence of point-cloud density changes. The PointNet series network has been widely used in the field of 3D object classification and point-cloud segmentation; however, when directly applied to facade parsing, its effect is far less than those of the other methods. The reason is that the special hierarchical structure and repeated layout of a facade (which directly determine the density of the point cloud) present the law of hierarchical and repeated distribution, while the scale and resolution of the PointNet++ network change uniformly globally, so it is unable to adaptively learn the structural characteristics of the facade.

#### Hybrid Methods

The SGPN method (Liu *et al.* 2019) integrates MVCNN and a graph network. The graph-network structure is constructed based on the original point-cloud spatial structure, with the advantage that point-cloud spatial-structure characteristics can be better learned theoretically. However, the shortcomings are also obvious. Even for small-scale facade point clouds, the directed acyclic graph network structure is very complex, the model parameters are huge, the computing power required is high, and the depth of the directed acyclic graph is not controlled, which may produce the phenomenon of gradient disappearance.

#### Combining the Two Categories

Knowledge-driven methods rely mainly on handcrafted design and expert experience; their feature-learning ability is not strong, and their algorithm robustness is not high. In contrast, methods based on deep learning have powerful feature-learning capabilities without the need to predefine complex rules, but they do not use the prior knowledge of the potential structure of the facade, which has a major effect on the performance of semantic segmentation. In this article, combining the advantages of prior knowledge and deep-learning methods, we propose a new method through generating an HMMV domain, which merges the edge prior information and texture information of the facade, and then using a multi-view

projection-based deep-learning method to extract more discriminative features, which can give a semantic segmentation model stronger generalizability. Our proposed strategy significantly improves the accuracy of 3D facade parsing in a point cloud.

## Methods

Through depth-information histogram statistics of facades, various elements in the facade point cloud show obvious hierarchical distribution characteristics, and the depth distribution of the same class of elements is consistent, which is the potential prior knowledge that can be incorporated into the model design. However, it is difficult to learn this feature by directly using multi-view convolutional neural networks on the original point cloud or mesh models. For this reason, this article designs an automatic method for extracting prior knowledge of the facade. A new submesh model

is constructed by fuzzy decomposition of the original mesh model. The mesh surface of the fuzzy area is eliminated to achieve coarse decomposition of the facade surface mesh. Therefore, on the mesh surface of the facade, there are no faces or edges between element patches (windows, balconies, walls, etc.), thus generating many spatially independent submeshes. These submeshes form a special mesh, which is called the H-Mesh in this article, as shown in Figure 3.

Before using 2D semantic segmentation models, an H-Mesh needs to be projected to the 2D multi-view domain, which is called the HMMV domain. The HMMV domain has a different data distribution compared with the original mesh multi-view (OMMV) domain. Intuitively, we propose that in the HMMV domain, the 2D semantic segmentation network can learn more discriminative features, thereby improving the accuracy of facade parsing. The comparative experiment in this article verified this intuition and insight. The overall process of the method is shown in Figure 1.

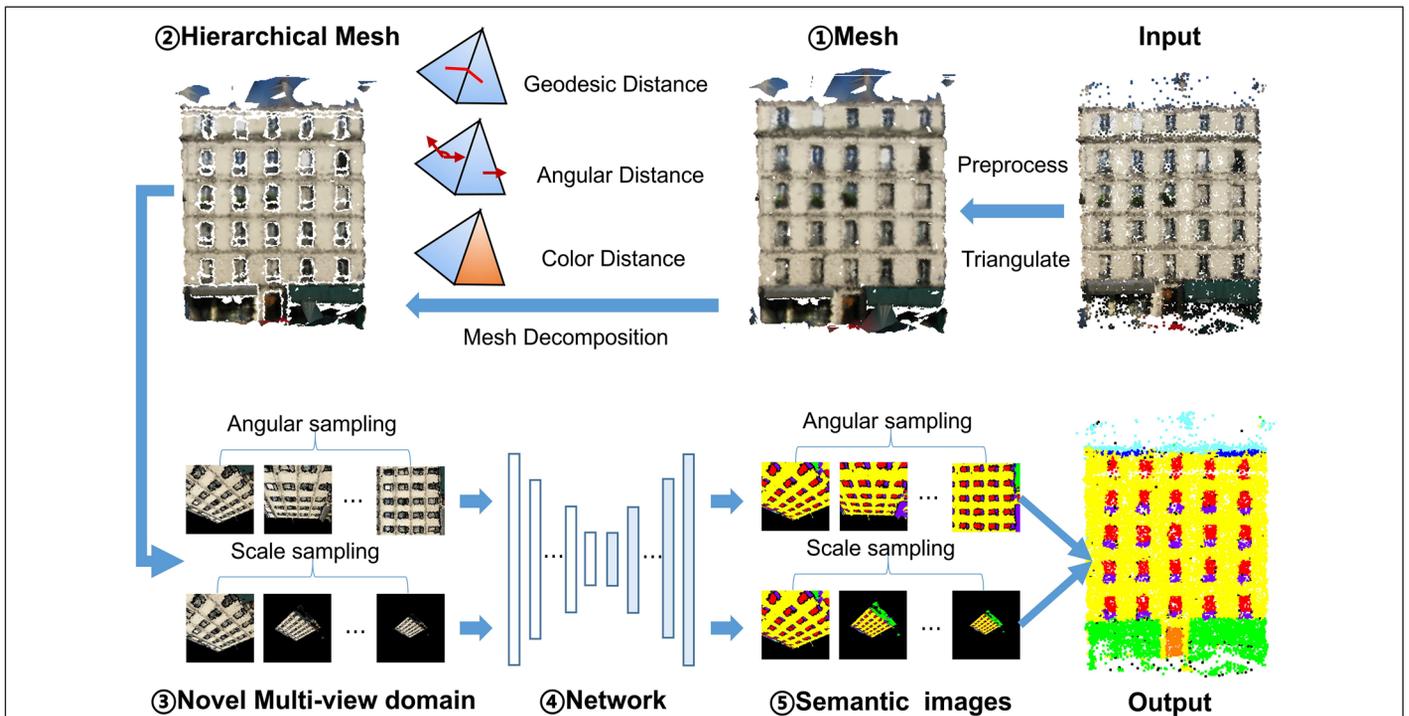


Figure 1. Overview of method. (1) Triangulated mesh generated from the original point cloud; (2) hierarchical mesh (H-Mesh) model generated based on the hierarchical mesh decomposition algorithm, which consists of multiple submeshes; (3) hierarchical mesh multi-view domain (HMMV) data generated by multi-scale and multi-angle 2D view sampling; (4) 2D segmentation network model—we used U-net and DeepLabv3+ networks separately to perform semantic segmentation on the HMMV data domain; (e5) the semantic segmentation results of HMMV, before projection back to the original point cloud to complete the 3D facade parsing.

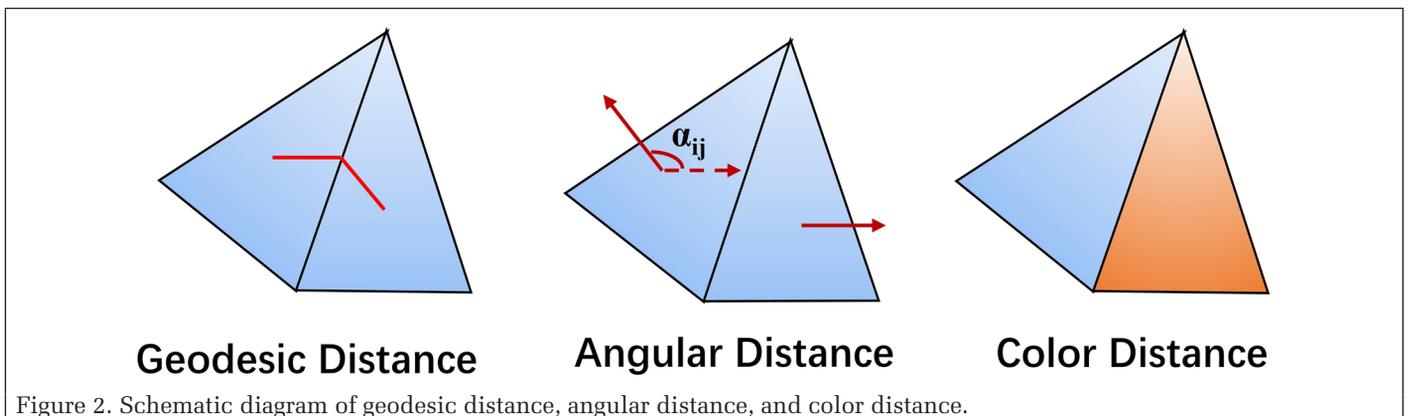


Figure 2. Schematic diagram of geodesic distance, angular distance, and color distance.

### Creation of the H-Mesh Model

The purpose of constructing the H-Mesh model is to extract the edge information of the facade elements. The distance feature vector is constructed through the geometric and color features between the various faces of the original mesh. The H-Mesh decomposition algorithm is used to obtain the coarse boundaries of elements, which are used as structural prior knowledge to improve semantic segmentation performance.

In the field of 3D geographic information systems and computer graphics, there are many mature standard modeling methods (Chew 1987; Blacker and Stephenson 1991; Owen 1998; Shewchuk 2002). In this article, we use the Delaunay triangulation-model method to generate the original mesh model, as shown in Figure 3. Next we decompose the original mesh model to generate an H-Mesh model. The purpose of mesh decomposition is to divide the mesh into several meaningful submeshes, but unlike with mesh segmentation, the submeshes formed by mesh decomposition—or fragments—are relatively coarse and do not contain any semantic label information. This process can be simply understood as a data-reconstruction process.

As the various elements of the facade mesh have obvious depth differences (as shown in Figure 3), the entire mesh surface is uneven. For example, windows are often sunken and balconies are generally convex. A hierarchical mesh decomposition (HMD) algorithm (Katz and Tal 2003) can decompose these uneven regions well. The key idea is to find meaningful components and at the same time create fuzzy boundaries between components. The algorithm is based on a basic assumption: that mesh faces that are far away from each other are unlikely to belong to the same submesh.

The process of HMD is mainly divided into two parts. First, the mesh is fuzzy clustered according to the geometric distance of the mesh surface, and then the fuzzy boundary area is optimized by the minimum-cut method. In the process of fuzzy clustering, the HMD algorithm considers only the two geometric features of geodesic distance and angular distance. However, the color distance between different types of element blocks in the facade mesh has obvious differences. Therefore, this article extends the color distance as one of the weights on the basis of the original algorithm, which effectively represents the texture difference of the facade

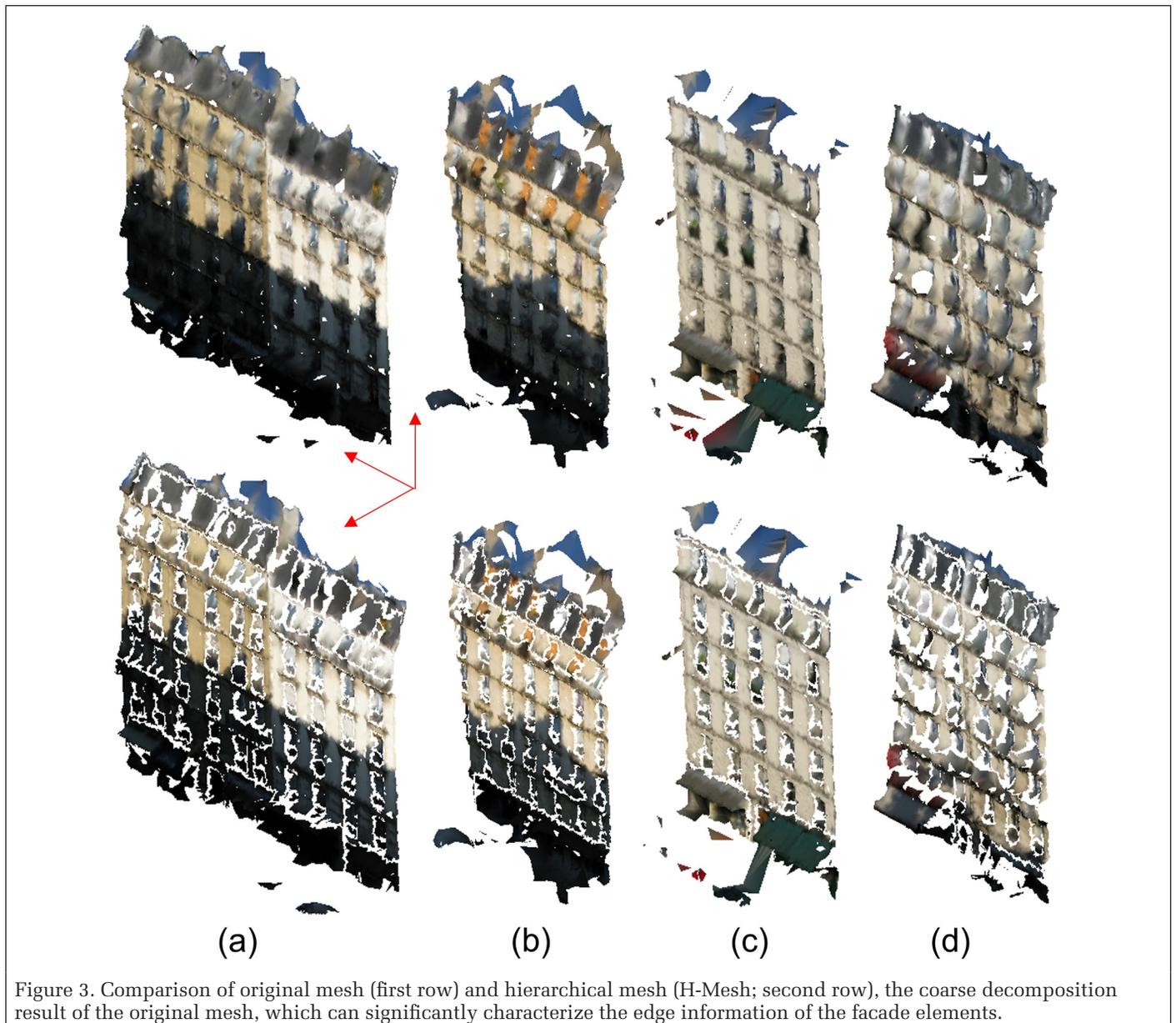


Figure 3. Comparison of original mesh (first row) and hierarchical mesh (H-Mesh; second row), the coarse decomposition result of the original mesh, which can significantly characterize the edge information of the facade elements.

elements and can further improve the mesh-decomposition performance. On the other hand, the fuzzy boundary areas between the components obtained by the original algorithm are intermediate results. This article eliminates these areas to form a new facade H-Mesh that contains many spatially separated submeshes.

In summary, the generation algorithm of the facade H-Mesh proposed in this article is mainly divided into three steps. The first step is to calculate the color distance and geometric distance (including geodesic distance and angular distance) and merge them to generate the edge weight of the dual mesh. The second step is to perform fuzzy clustering using the weight of the mesh surface to complete the preliminary decomposition of the elements of the facade and generate the fuzzy area. In the third step, the fuzzy area of mesh is removed; the remaining parts form the H-Mesh of the facade with coarse element boundaries. The fuzzy process in the second step is consistent with the HMD method (Katz and Tal 2003), so we do not repeat it here. Next we elaborate on the first and third steps.

#### Fusing Color Distance and Geometric Distance

As the RueMonge data set was collected under natural light, the color distance in Equation 1 is used in this article to be closer to the  $L^*u^*v^*$  uniform color space; that is, the color difference represented by the spatial distance is consistent with the difference in human vision. Thus, the chromatic aberration between two grid surfaces can be judged more realistically and stably. Given two adjacent grid surfaces  $f_i$  and  $f_j$ , their original color space is  $(R, G, B)$ —the three primary colors—and the color distance is defined as

$$\text{Color\_Dist}(f_i, f_j) = \sqrt{\left(2 + \frac{\bar{r}}{256}\right)\Delta R^2 + 4\Delta G^2 + \left(2 + \frac{255 - \bar{r}}{256}\right)\Delta B^2}, \quad (1)$$

where

$$\bar{r} = \frac{R_i + R_j}{2}, \quad \Delta R = R_i - R_j, \quad \Delta G = G_i - G_j, \quad \Delta B = B_i - B_j. \quad (2)$$

$\text{Geo\_Dist}(f_i, f_j)$  is the shortest distance along the mesh surface between the geometric center points of the mesh faces  $f_i$  and  $f_j$ ;

$$\text{Geo\_Dist}(f_i, f_j) = \text{Shortest\_Distance}(\text{Center}_{f_i}, \text{Center}_{f_j}). \quad (3)$$

$\text{Ang\_Dist}(f_i, f_j)$  is the angular distance between the normal vectors of the mesh faces,

$$\text{Ang\_Dist}(f_i, f_j) = \eta(1 - \cos \alpha_{ij}), \quad (4)$$

where  $\alpha_{ij}$  is the angle between the normal vectors of the mesh faces  $f_i$  and  $f_j$ , and  $\eta$  is the factor for the convexity of the mesh face. The color distance between adjacent mesh faces characterizes the texture difference of the facade elements, whereas the angular distance and geodesic distance characterize the geometric structure difference of the facade elements at the boundary; therefore, they are merged to generate an edge weight of the dual mesh, which can more comprehensively reflect the features of the mesh faces, generate a more accurate shortest path, and obtain better fuzzy decomposition results. The edge weight of the dual mesh is

$$\text{Weight}(f_i, f_j) = \delta_1 \frac{\text{Geo\_Dist}(f_i, f_j)}{\text{avg}(\text{Geo\_Dist})} + \delta_2 \frac{\text{Ang\_Dist}(f_i, f_j)}{\text{avg}(\text{Ang\_Dist})} + (1 - \delta_1 - \delta_2) \frac{\text{Color\_Dist}(f_i, f_j)}{\text{avg}(\text{Color\_Dist})}, \quad (5)$$

where  $\text{avg}(\text{Color\_Dist})$  represents the average color distance of all adjacent faces in the mesh,  $\text{avg}(\text{Geo\_Dist})$  represents the average geodesic distance of the center points of all adjacent faces in the mesh, and  $\text{avg}(\text{Ang\_Dist})$  represents the average angular distance of all adjacent faces in the mesh.

#### Removing Fuzzy Areas

After the first two steps are completed, the original mesh is divided into many independent mesh patches, which are defined as a set  $p$ ; the  $j$ th patch is  $p_j$ , and the probability that the face of mesh  $f_i$  belongs to patch  $p_j$  is  $P(f_i, p_j)$ . The distribution is

$$\text{fuzzy} = \begin{cases} f_i \in p_j, \notin \text{fuzzy}, & P(f_i, p_j) > 0.5 + \varepsilon \\ f_i \in \text{fuzzy}, & 0.5 - \varepsilon \leq P(f_i, p_j) \leq 0.5 + \varepsilon. \\ f_i \notin p_j, \notin \text{fuzzy}, & P(f_i, p_j) < 0.5 - \varepsilon \end{cases} \quad (6)$$

In this step, we traverse the entire mesh and remove the faces with probability values in the interval  $[0.5 - \varepsilon, 0.5 + \varepsilon]$  as fuzzy areas, which completes the mesh separation of adjacent facade elements. Removing the fuzzy area loses only a small part of the vertex data; however, the generated H-Mesh significantly represents the prior knowledge of the facade spatial layout and element boundary information, which is precisely the key to improving the accuracy of the facade parsing.

In Figure 3, the first row is the original mesh and the second row is the H-Mesh. When the point-cloud data of the facade are relatively complete and the density is appropriate, based on the H-Mesh decomposition method, a higher-quality facade H-Mesh can be generated, as shown in Figure 3a–3c. When facade point-cloud data are missing, such as in Figure 3d, the decomposition performance will inevitably be affected by the distortion of the distance between the local mesh surfaces.

#### Generation of the HMMV Data Domain

The HMMV data domain is generated to express the 3D H-Mesh completely through 2D projection. The HMMV data combine the prior edge information and texture information of the facade and are the input of the 2D semantic segmentation model.

As the internal submesh of the H-Mesh presents a discrete state in the spatial topological relationship and cannot directly perform feature learning on the surface of the mesh, this article uses 2D projection to generate an HMMV data domain, which not only has rich texture features but also has significant edge features. The data distribution is different from in the multi-view data domain generated by the original mesh. This contains the constraints of structure prior knowledge, so as to improve the feature-description ability of deep neural networks.

Inspired by MVCNN and SnapNet (Su *et al.* 2015; Boulch *et al.* 2018), according to the characteristics of the facade object we constructed a new virtual camera that can sample the 2D image of the H-Mesh from multiple angles and multiple scales. This sampling strategy can not only fully express the facade objects but also help with the deep-network learning of salient features, as well as indirectly realize data augmentation, preventing overfitting in later model training.

The virtual-camera model is located on the surface of a 3D sphere. The center of the sphere is the geometric center point of the elevation point cloud. The sampling direction is determined by the normal vector of the mesh surface. Through setting equal-spaced angles, multi-view image sampling is completed. Setting a different sphere radius in each angle achieves multi-scale sampling. The specific process is shown in Figure 4.

By projecting the data onto a 2D plane, a new HMMV domain is generated, in which various elements of the facade are discrete from each other in topological space. This can help humans intuitively perform visual interpretation. Our observation also shows that data distribution in the HMMV domain is more consistent with the laws of human vision, and that the deep neural network in the HMMV domain can extract more discriminative features.

#### Semantic Segmentation and Back Projection

The input of the semantic segmentation model is the HMMV data, and the output is multi-view images with semantic

labels. The input of the back-projection stage is semantically labeled multi-view images, and the output is a facade point cloud with semantic labels.

Generally, the image semantic segmentation network is mainly composed of an encoder and a decoder. The encoder extracts features from the original image, and the decoder up-samples the feature map generated at the encoding stage to restore the semantic information. This article was based on the U-net (Ronneberger *et al.* 2015) and DeepLabv3+ (Chen *et al.* 2018) semantic segmentation models. The encoder uses VGG and ResNet convolutional neural networks (Simonyan and Zisserman 2015; He *et al.* 2016), respectively, to verify the robustness and universality of the HMMV domain to different network structures.

After completing all the 2D HMMV image semantic segmentation, according to the virtual-camera model already defined, the 2D semantic labels are back-projected to the 3D point cloud. The back-projection process is similar to SnapNet (Boulch *et al.* 2018), and the essence is integrating predicted values from different perspectives. Specifically, since the vertices of the mesh and the pixels of the view are in a one-to-many mapping relationship, the same mesh vertices will appear in multiple views. According to the projected relationship defined in the sampling stage, the predicted value is assigned to the corresponding mesh vertex in the reverse direction. Through iterating through all the views, the label of the mesh vertex is the category corresponding to the maximum predicted value. Finally, through the KD tree index generated in the initial mesh-model creation stage, the nearest mesh vertex  $v$  of each point  $p$  in the facade point cloud is obtained, and the label of this vertex  $v$  is the label of point  $p$ , thus completing the semantic segmentation of the facade point cloud.

## Experiments

Since there are few publicly available semantic segmentation data sets for 3D facades, we mainly conducted experiments on the RueMonge 2014 data set (Riemenschneider *et al.* 2014) to test the performance of the method based on the HMMV data domain proposed in this article. The experimental results showed that the method substantially improved the accuracy of 3D facade parsing. Our method was run on the PyTorch version 1.1.0 (Paszke *et al.* 2019) deep-learning platform. All experiments were performed in the same hardware environment equipped with an Intel 7700K central processing unit and Nvidia GTX 1080Ti graphics processing unit.

### Data Set

The ETHZ CVL RueMonge 2014 building-facade data set (Riemenschneider *et al.* 2014) is used for the 3D reconstruction of the facade in the urban scene and the semantic segmentation of the facade point cloud. The data set is sparse point-cloud data and was created based on the SfM reconstruction and the multi-view stereo method. The architectural design style is Haussmannian. So far, RueMonge 2014 is the largest and only public data set in the field of 3D facade segmentation. The scene of this data set is the Rue Monge in Paris, France, with a total length of 700 m. The facade point clouds along the street are all point-wise semantically labeled. A total of seven semantic categories are labeled: windows, walls, balconies, doors, roofs, sky, and shops. Access to the RueMonge 2014 3D facade data set can be requested at <http://www.varcity.eu/3dchallenge/>.

### Model Specification

On the RueMonge 2014 facade-parsing task, we adopted three main evaluation measures to evaluate our method against state-of-the-art methods: overall pixel accuracy, average class accuracy, and mean intersection over union (IoU). The mean IoU represents the average of the ratio of the intersection and union between the true and predicted values of each element

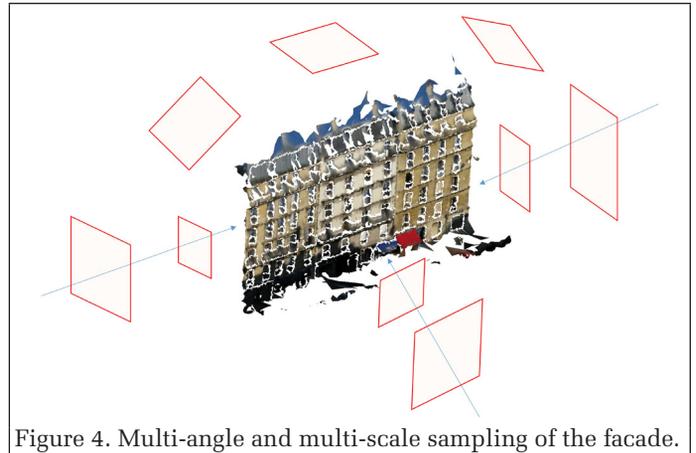


Figure 4. Multi-angle and multi-scale sampling of the facade.

of the facade, and is the most important indicator for measuring the performance of the 3D facade-parsing algorithm.

We designed four comparative experiments to verify the performance of our methods. As shown in Tables 1 and 2, the semantic segmentation model used U-net and DeepLabv3+, respectively, and the input data was the HMMV and OMMV data. The multi-views were all images of size  $224 \times 224$  pixels, the multi-angle sampling range was  $(0^\circ, 180^\circ)$ , and the multi-scale sampling parameters were  $[3, 5, 7, 10]$ ; thus, there were four sampling distances in total. In the experiments, we set a total of 17 sampling angles, and therefore the collection size of the multi-view data domain of a single facade mesh was  $17 \times 4$ . The encoder adopted the VGG-16 and ResNet-34 network structures, and used ImageNet pretrained weights as the model initialization parameters. An Adam optimizer was used, and the initial learning rate was 0.0001. All models were trained in parallel on two GTX 1080Ti graphics processing units, and the parameters of the model were updated 5100 times.

### Results on the RueMonge 2014 Data Set

#### Quantitative Analysis

To evaluate the parsing performance of our methods in detail, we quantitatively analyzed (Table 1) the four proposed methods (Ours1–4) and the current six main methods—OctNet (Riegler *et al.* 2017), 3D All the Way (Martinovi *et al.* 2015), Auto-Context (Gadde *et al.* 2018), SPLATNet3D (Su *et al.* 2018), SPLATNet-2D&3D (Su *et al.* 2018), and SGPN (Liu *et al.* 2019)—on the RueMonge 2014 data set. Our proposed semantic segmentation method (Ours4) based on the HMMV data domain and DeepLabv3+ achieved 76.41% on the core indicator of the mean IoU, leading the current best result by 2.75%. The other two indicators were also on par with the current best results. In comparing the experimental results of Ours1–4 using the same semantic segmentation model, the HMMV-based method was significantly better than the OMMV-based method, with respective increases of 6.53% and 6.93%. This is a significant improvement, indicating that the performance based on the HMMV method is superior.

#### Qualitative Analysis

In this section, we compare only the published prediction results. A visual qualitative analysis of the facade-parsing results was performed to intuitively compare the advantages of the proposed method against the other methods, as shown in Figure 5. SPLATNet-3D is a SplatNet method that uses only 3D features, Ours3 is a method based on OMMV, and Ours4 is a method based on HMMV. These use DeepLabv3+ as the semantic segmentation model. Figure 5 shows the test results of each method on the RueMonge 2014 facade-segmentation data set. Using the ground truth as a reference, we can intuitively see that the HMMV-based method (Ours4) performed better than the other methods.

Specifically, it is difficult for human eyes to recognize windows that are blocked by shadows (the first row) and shops with complex structures (the second row). Benefiting from the prior knowledge of the fuzzy boundary of the H-Mesh, the

Table 1. Three main evaluation indicators of different methods.

Method	Average <sup>a</sup>	Overall <sup>b</sup>	Mean IoU
OctNet (Riegler <i>et al.</i> 2017)	73.60	81.50	59.20
3D All the Way (Martinović <i>et al.</i> 2015)	—	—	61.39
Auto-Context (Gadde <i>et al.</i> 2018)	75.30	84.70	62.90
SPLATNet3D (Su <i>et al.</i> 2018)	79.34	84.62	65.40
SPLATNet-2D&3D (Su <i>et al.</i> 2018)	—	—	69.80
SGPN (Liu <i>et al.</i> 2019)	—	—	73.66
Ours1 (OMMV with U-net)	76.62	88.52	66.75
Ours2 (HMMV with U-net)	82.18	90.21	73.28
Ours3 (OMMV with DeepLabv3+)	79.57	89.17	69.48
<b>Ours4 (HMMV with DeepLabv3+)<sup>c</sup></b>	<b>84.55</b>	<b>91.15</b>	<b>76.41</b>

HMMV = hierarchical mesh multi-view domain; IoU = intersection over union; OMMV = original mesh multi-view domain.

<sup>a</sup>Overall pixel accuracy.

<sup>b</sup>Average class accuracy.

<sup>c</sup>Boldface indicates the best segmentation results of all facade elements.

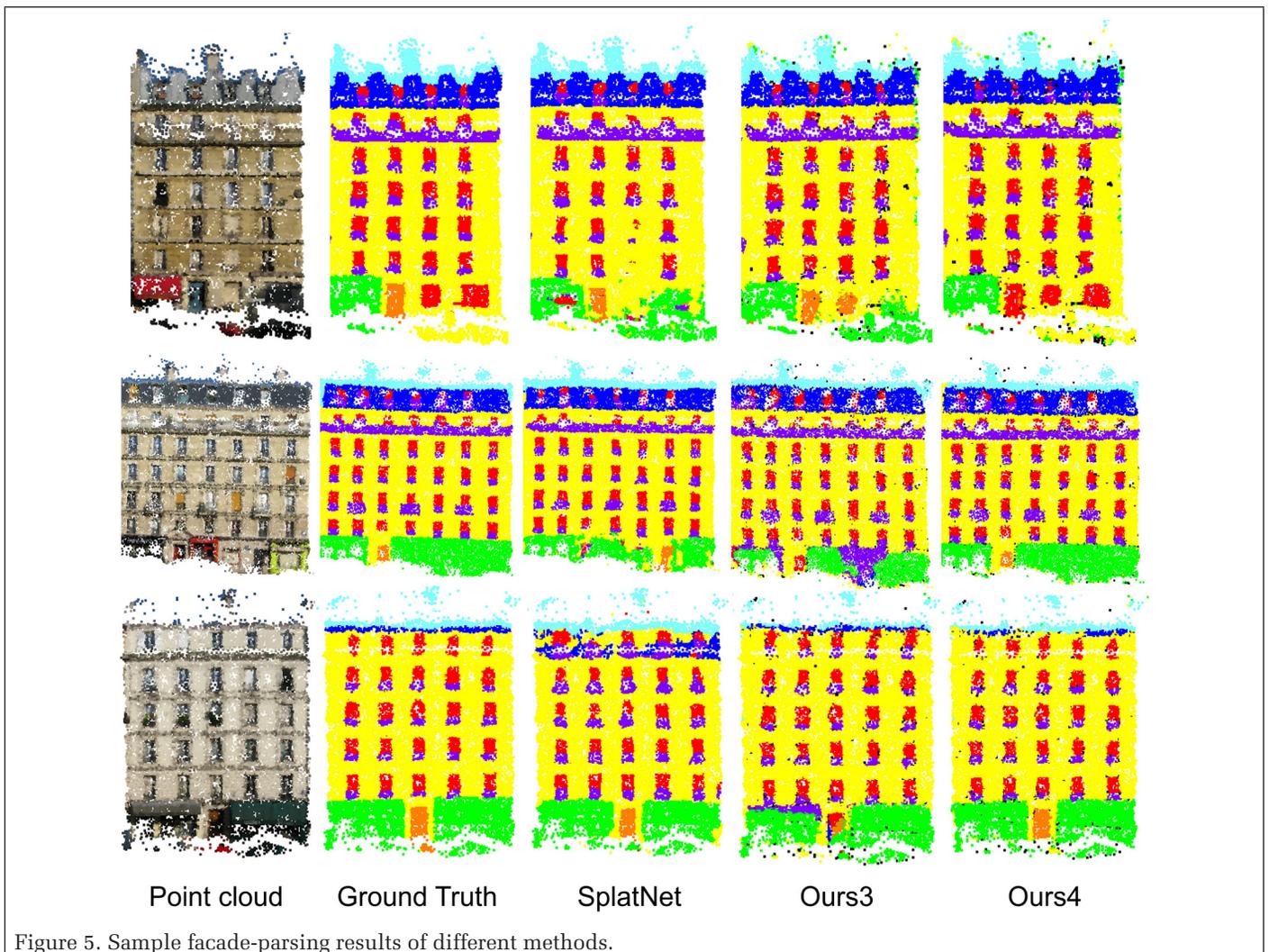
HMMV = hierarchical mesh multi-view domain; IoU = intersection over union; OMMV = original mesh multi-view domain.

Boldface indicates the best segmentation results of all facade elements.

Ours4 method can identify facade elements completely, while the segmentation results of the other two methods are not ideal. In addition, for the situation of the facade without occlusion and shadow (the third row), SPLATNet3D, due to the lack of learning the 2D texture features, divided part of the wall into balconies. The method based on OMMV (Ours3), due to the lack of 3D geometric features, mistakenly divided shops with similar colors into balconies, but the Ours4 method accurately completed the parsing of the entire facade, showing strong robustness.

## Discussion

To explore the reasons for the improved performance of the method proposed in this article, we combined two types of input data and two semantic segmentation networks in pairs, and designed four sets of controlled experiments. We adopted the controlled variable method for comparative experiments; that is, only the input mesh multi-view data were changed, and the other factors remained unchanged. First, multi-view sampling was performed on the original mesh, and the virtual-camera model was kept consistent. Then the same multi-view sampling was performed on the H-Mesh, and the model training was performed without any other data set or data augmentation. The hyper-parameters of the deep-segmentation network were consistent. Finally, we evaluated the prediction results based on the official evaluation code. We analyze the results in detail.



From Table 2, based on the proposed facade-parsing method of the HMMV, the segmentation accuracy of most facade elements was significantly improved, especially for doors and shops. Normally, these two types of elements are distributed at the bottom of the facade. During data collection, the shadow of the sun had a greater impact on the texture expression of the surface. Therefore, the method based on OMMV cannot learn discriminative texture features to recognize facade elements, whereas the HMMV-based method incorporates geometric edge information, so that the encoder of the semantic segmentation model can focus more on the area within the edge.

The performance for walls and roofs did not improve significantly. The main reason is that both of them are large-area

plane elements in the facade structure, and the concave/convex properties are not obvious; thus, the geometry feature descriptions in the H-Mesh and the original mesh are relatively consistent, and as a result the parsing accuracies in the same semantic segmentation model are not greatly different. Windows and balconies, on the other hand, are concave and convex structures; they are clearly separated from other facade elements in the HMMV data domain, as shown in Figure 3, and so the generated feature maps are more compact during the encoding stage of the semantic segmentation model, thus improving parsing performance.

To verify the generation ability of the HMMV-based method, we conducted an experiment with cross-domain input data, taking OMMV data as the input and using the model trained

Table 2. Semantic segmentation results for different categories of facade elements.<sup>a</sup>

	Input Data	Model	Class Mean IoU	Window	Wall	Balcony	Door	Roof	Sky	Shop
Ours1	OMMV	U-net	66.75	65.56	88.96	68.01	23.66	72.08	73.93	75.05
Ours2	HMMV	U-net	73.28	71.45	88.63	75.23	45.55	<b>72.25</b>	77.32	82.57
Ours3	OMMV	DeepLabv3+	69.48	71.49	88.84	72.70	39.74	71.76	66.33	75.48
Ours4	HMMV	DeepLabv3+	<b>76.41</b>	<b>76.81</b>	<b>89.10</b>	<b>78.78</b>	<b>54.40</b>	70.74	<b>81.37</b>	<b>83.68</b>

HMMV = hierarchical mesh multi-view domain; IoU = intersection over union; OMMV = original mesh multi-view domain.

<sup>a</sup>Boldface indicates the best segmentation results for each facade element,

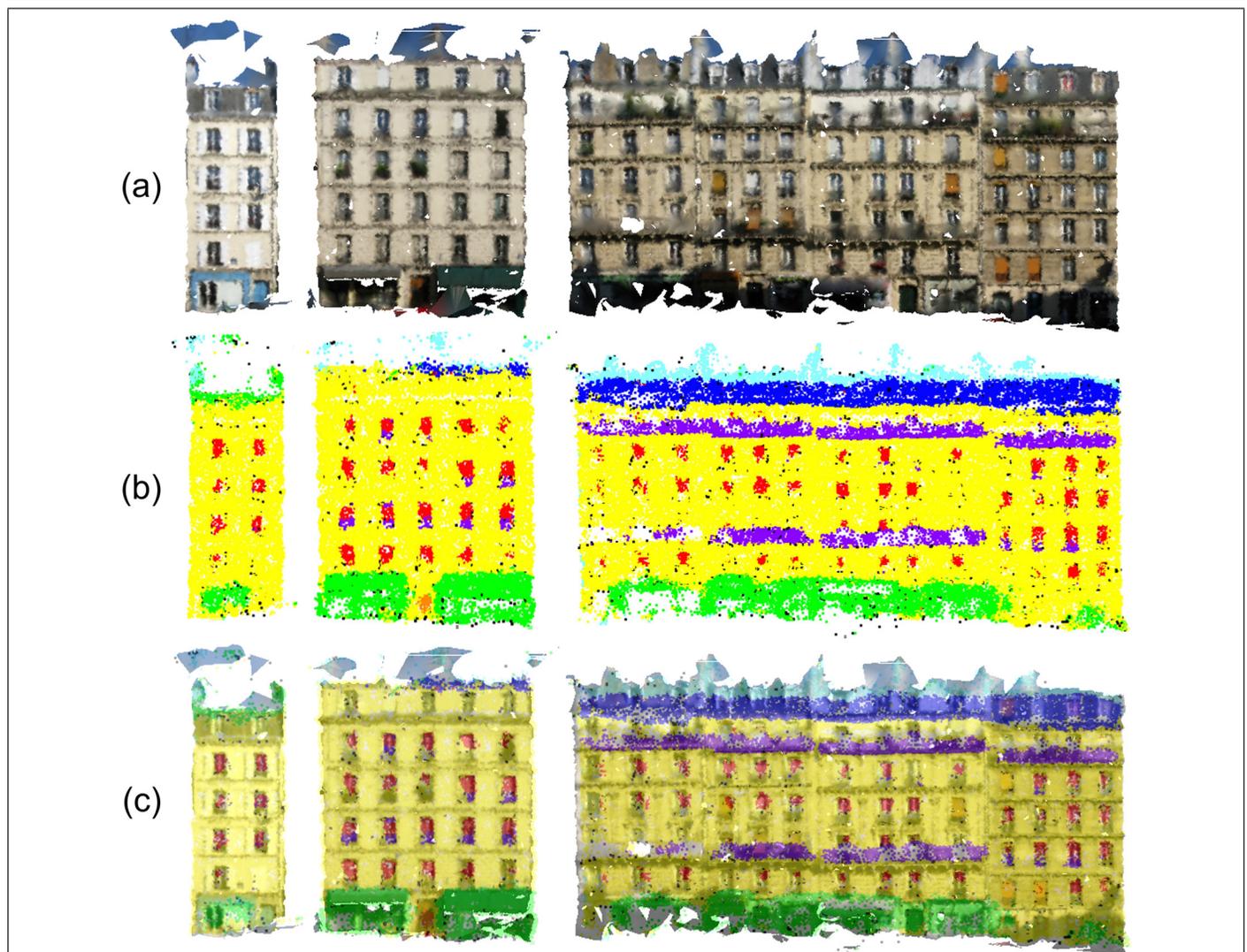


Figure 6. Prediction of the original mesh multi-view data through the semantic segmentation model trained on the hierarchical mesh multi-view data domain. (a) The original mesh; (b) the cross-domain facade-parsing results; (c) the segmentation mask on the original mesh surface.

in the HMMV data domain to perform semantic segmentation. From the segmentation results (shown in Figure 6b), the method based on the HMMV data domain learned that the obtained features had a more essential description of the data, and also had certain semantic segmentation capabilities for the input of the original mesh, which was cross-domain.

Figure 7a is the input data from the HMMV domain, and Figure 7b is the input data from the OMMV domain. The “holes” at the edges of H-Mesh are mapped as background information. Feature maps Conv1\_6 and Conv1\_34 represent the filter or channel generated in the first convolution layer of the encode stage of the semantic segmentation model. Feature map Conv1\_6 shows the feature extracted by the encoder on the window boundary of the facade, and Feature map Conv1\_34 shows the feature extracted by the encoder on the wall area. In comparison between them, based on the H-Mesh the encoder could better characterize the features, thereby improving the semantic segmentation performance. By comparing the results of the four experiments, we can draw the following conclusions: the HMMV data domain was invariant to the network structure, and in the HMMV data domain, the model was able to extract more salient features of the facade elements.

Although the proposed method can be applied to most types of urban facades, it still has limitations. For facade types that do not conform to the characteristics of the hierarchical layout, such as office blocks or towering hotels, their facades are often composed of a whole glass plane, and the depth information of various elements is similar. It is difficult to extract boundary features of the elements using our method. At present, the proposed method is based on the multi-view deep-learning model, which is more complicated. The next step is to use the deep-network model to directly perform semantic segmentation on 3D facade mesh to form a simple and efficient end-to-end 3D facade-parsing method. Since there are few public 3D point-cloud facade data sets

currently, we only conduct experiments with the RueMonge 2014 3D facade data set. In the future, we will use visual-reconstruction technology to generate a 3D facade data set of Chinese street scenes based on Internet street-view map data. We hope that it will attract more researchers to participate in the work in this field, and will be beneficial for verifying the robustness of the algorithm. In addition, due to the high cost of creating a 3D facade data set and the variety of facade styles, conducting cross-domain learning based on a small sample of a facade data set is a research hot spot in the future of 3D facade parsing.

## Conclusions

The new data domain constructed based on the hierarchical mesh is essentially an initial feature map, which contains fuzzy object-boundary information, similar to the feature map output of the shallow structure in the deep-network model (often including object boundaries, color features, etc.); however, the process of feature extraction is completely different, the output results of the convolutional neural network model are poor in interpretability, and feature extractors often use local information of the image. In contrast, the method in this article uses the geometric distance and color distance of the mesh to construct features, and these features have real physical meanings, comport with human visual laws, and make full use of the global information of the facade during the fuzzy clustering process. Thus, the decomposition results are highly reliable and can guide the encoder to extract more salient features.

At present, in the field of computer vision perception, the advancement of the technology has entered a bottleneck period with the development of the basic model of deep learning and improvement of its accuracy. The performance improvements brought by designing more complex network structures

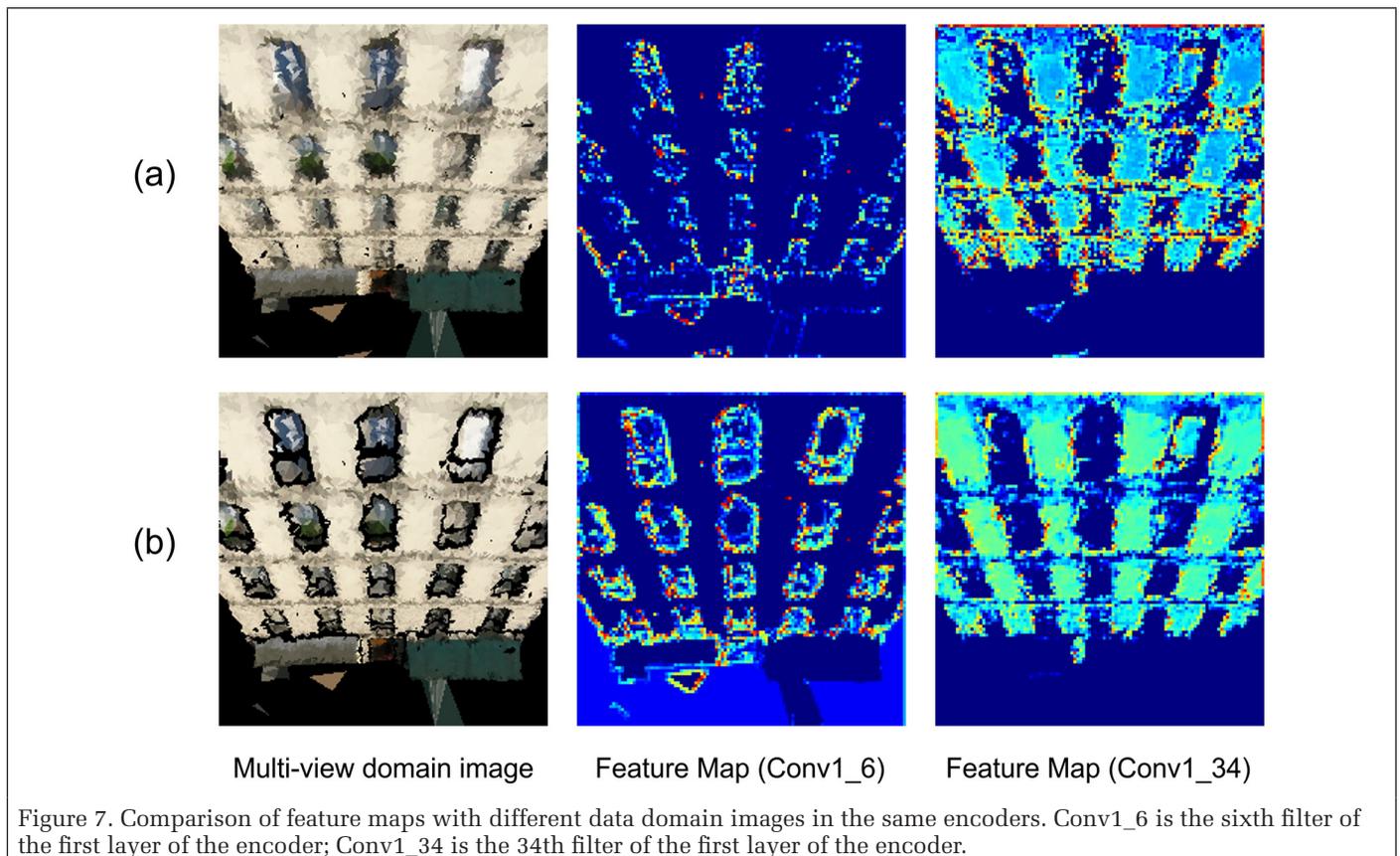


Figure 7. Comparison of feature maps with different data domain images in the same encoders. Conv1\_6 is the sixth filter of the first layer of the encoder; Conv1\_34 is the 34th filter of the first layer of the encoder.

are limited and consume huge computing resources. From the perspective of data distribution, in this article we constructed a new hierarchical-mesh multi-view data domain based on the characteristics of a facade object to achieve fusion of deep-learning models and prior knowledge, thereby significantly improving the segmentation accuracy. We believe that this research idea will also show value in other applications.

## Acknowledgments

This research is funded by the National Key Research and Development Program of China (project 2018YFB0505400)

## References

- Anguelov, D., B. Taskarf, V. Chatalbashev, D. Koller, D. Gupta, G. Heitz and A. Ng. 2005. Discriminative learning of Markov random fields for segmentation of 3D scan data. Pages 169–176 in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 2, held in San Diego, Calif., 20–25 June 2005. Edited by J. Editor. City, St.: IEEE.
- Biljecki, F., H. Ledoux and J. Stoter. 2016. An improved LOD specification for 3D building models. *Computers, Environment and Urban Systems* 59:25–37.
- Biljecki, F., H. Ledoux, J. Stoter and J. Zhao. 2014. Formalisation of the level of detail in 3D city modelling. *Computers, Environment and Urban Systems* 48:1–15.
- Blacker, T. D. and M. B. Stephenson. 1991. Paving: A new approach to automated quadrilateral mesh generation. *International Journal for Numerical Methods in Engineering* 32 (4):811–847.
- Boulch, A., J. Guerry, B. Le Saux and N. Audebert. 2018. SnapNet: 3D point cloud semantic labeling with 2D deep segmentation networks. *Computers & Graphics* 71:189–198.
- Charles, R. Q., H. Su, M. Kaichun and L. J. Guibas. 2017. PointNet: Deep learning on point sets for 3D classification and segmentation. Pages 77–85 in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, held in Honolulu, Hawaii, 21–26 July 2017. Edited by J. Editor. City, St.: IEEE.
- Chen, L.-C., Y. Zhu, G. Papandreou, F. Schroff and H. Adam. 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. Pages 833–851 in *Computer Vision—ECCV 2018*, held in Munich, Germany, 8–14 September 2018. Edited by V. Ferrari, M. Hebert, C. Sminchisescu and Y. Weiss. Lecture Notes in Computer Science vol. 11214. Cham, Switzerland: Springer.
- Chew, L. P. 1987. Constrained Delaunay triangulations. Pages 215–222 in *SCG '87: Proceedings of the Third Annual Symposium on Computational Geometry*, held in Waterloo, Ontario, June 1987. Edited by D. Soule. New York: Association for Computing Machinery.
- Chiang, H.-Y., Y.-L. Lin, Y.-C. Liu and W. H. Hsu. 2019. A unified point-based framework for 3D segmentation. Pages 155–163 in *2019 International Conference on 3D Vision (3DV)*, held in Québec City, Quebec, 16–19 September 2019. Edited by J. Editor. City, St.: IEEE.
- Dai, A. and M. Nießner. 2018. 3DMV: Joint 3D-multi-view prediction for 3D semantic scene segmentation. Pages 458–474 in *Computer Vision—ECCV 2018*, held in Munich, Germany, 8–14 September 2018. Edited by V. Ferrari, M. Hebert, C. Sminchisescu and Y. Weiss. Lecture Notes in Computer Science vol. 11214. Cham, Switzerland: Springer.
- Gadde, R., V. Jampani, R. Marlet and P. V. Gehler. 2018. Efficient 2D and 3D facade segmentation using auto-context. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40 (5):1273–1280.
- Gröger, G. and L. Plümer. 2012. CityGML—interoperable semantic 3D city models. *ISPRS Journal of Photogrammetry and Remote Sensing* 71:12–33.
- He, K., X. Zhang, S. Ren and J. Sun. 2016. Deep residual learning for image recognition. Pages 770–778 in *2016 IEEE Conference on Computer Vision and Pattern Recognition*, held in Las Vegas, Nev., 27–30 June 2016. Edited by J. Editor. City, St.: IEEE.
- Jaritz, M., J. Gu and H. Su. 2019. Multi-View PointNet for 3D scene understanding. Pages 3995–4003 in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, held in Seoul, South Korea, 27–28 October 2019. Edited by J. Editor. City, St.: IEEE.
- Jing, H. and S. You. 2016. Point cloud labeling using 3D Convolutional Neural Network. Pages 2670–2675 in *2016 23rd International Conference on Pattern Recognition*, held in Cancun, Mexico, 4–8 December 2016. Edited by J. Editor. City, St.: IEEE.
- Katz, S. and A. Tal. 2003. Hierarchical mesh decomposition using fuzzy clustering and cuts. Pages 954–961 in *SIGGRAPH '03: ACM SIGGRAPH 2003 Papers*, held in San Diego, Calif., July 2003. Edited by J. Editor. New York: Association for Computing Machinery.
- Kozi ski, M., R. Gadde, S. Zagoruyko, G. Obozinski and R. Marlet. 2015. A MRF shape prior for facade parsing with occlusions. Pages 2820–2828 in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, held in Boston, Mass., 7–12 June 2015. Edited by J. Editor. City, St.: IEEE.
- Ladický, L., P. Sturgess, C. Russell, S. Sengupta, Y. Bastanlar, W. Clocksin and P.H.S. Torr. 2012. Joint optimization for object class segmentation and dense stereo reconstruction. *International Journal of Computer Vision* 100:122–133.
- Lawin, F. J., M. Danelljan, P. Tosteberg, G. Bhat, F. S. Khan and M. Felsberg. 2017. Deep projective 3D semantic segmentation. Pages 95–107 in *Computer Analysis of Images and Patterns*, held in Ystad, Sweden, 22–24 August 2017. Edited by M. Felsberg, A. Heyden and N. Krüger. Lecture Notes in Computer Science vol. 10424. Cham, Switzerland: Springer.
- Li, Y., Q. Zheng, A. Sharf, D. Cohen-Or, B. Chen and N. J. Mitra. 2011. 2D-3D fusion for layer decomposition of urban facades. Pages 882–889 in *2011 International Conference on Computer Vision*, held in Barcelona, Spain, 6–13 November 2011. Edited by J. Editor. City, St.: IEEE.
- Li, Z., L. Zhang, P. T. Mathiopoulos, F. Liu, L. Zhang, S. Li and H. Liu. 2017. A hierarchical methodology for urban facade parsing from TLS point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing* 123:75–93.
- Lin, T.-Y., P. Dollár, R. Girshick, K. He, B. Hariharan and S. Belongie. 2017. Feature pyramid networks for object detection. Pages 936–944 in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, held in Honolulu, Hawaii, 21–26 July 2017. Edited by J. Editor. City, St.: IEEE.
- Liu, S., X. Li, V. Jampani, S. Mello and J. Kautz. 2019. Learning propagation for arbitrarily-structured data. Pages 652–661 in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, held in Seoul, South Korea, 27 October–2 November 2019. Edited by J. Editor. City, St.: IEEE.
- Martinovi, A., J. Knopp, H. Riemenschneider and L. Van Gool. 2015. 3D all the way: Semantic segmentation of urban scenes from start to end in 3D. Pages 4456–4465 in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, held in Boston, Mass., 7–12 June 2015. Edited by J. Editor. City, St.: IEEE.
- Martinovic, A. and L. Van Gool. 2013. Bayesian grammar learning for inverse procedural modeling. Pages 201–208 in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, held in Portland, Ore., 23–28 June 2013. Edited by J. Editor. City, St.: IEEE.
- Meng, H.-Y., L. Gao, Y.-K. Lai and D. Manocha. 2019. VV-net: Voxel VAE net with group convolutions for point cloud segmentation. Pages 8499–8507 in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, held in Seoul, South Korea, 27 October–2 November 2019. Edited by J. Editor. City, St.: IEEE.

- Munoz, D., J. A. Bagnell and M. Hebert. 2012. Co-inference for multi-modal scene analysis. Pages 668–681 in *Computer Vision—ECCV 2012*, held in Florence, Italy, 7–13 October 2012. Edited by A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato and C. Schmid. Lecture Notes in Computer Science vol. 7577. Berlin: Springer.
- Owen, S. J. 1998. A survey of unstructured mesh generation technology. Pages 239–267 in *Proceedings, 7th International Meshing Roundtable*, held in Dearborn, Mich., 26–28 October 1998. Edited by J. Editor. City, St.: Publisher.
- Paszke, A., S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai and S. Chintala. 2019. PyTorch: An imperative style, high-performance deep learning library. Pages 8026–8037 in *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)*, held in City, St., DD–DD Month YYYY. Edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox and R. Garnett. City, St.: Curran Associates.
- Qi, C. R., L. Yi, H. Su and L. J. Guibas. 2017. PointNet++: Deep hierarchical feature learning on point sets in a metric space. Pages 5105–5114 in *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, held in City, St., DD–DD Month YYYY. Edited by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett. City, St.: Publisher.
- Riegler, G., A. O. Ulusoy and A. Geiger. 2017. OctNet: Learning deep 3D representations at high resolutions. Pages 6620–6629 in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, held in Honolulu, Hawaii, 21–26 July 2017. Edited by J. Editor. City, St.: Publisher.
- Riemenschneider, H., A. Bódis-Szomorú, J. Weissenberg and L. Van Gool. 2014. Learning where to classify in multi-view semantic segmentation. Pages 516–532 in *Computer Vision—ECCV 2014*, held in Zurich, Switzerland, 6–12 September 2014. Edited by D. Fleet, T. Pajdla, B. Schiele and T. Tuytelaars. Lecture Notes in Computer Science vol. 8693. Cham, Switzerland: Springer.
- Riemenschneider, H., U. Krispel, W. Thaller, M. Donoser, S. Havemann, D. Fellner and H. Bischof. 2012. Irregular lattices for complex shape grammar facade parsing. Pages 1640–1647 in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, held in Providence, R.I., 16–21 June 2012. Edited by J. Editor. City, St.: IEEE.
- Ronneberger, O., P. Fischer and T. Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. Pages 234–241 in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*, held in Munich, Germany, 5–9 October 2015. Edited by N. Navab, J. Hornegger, W. Wells and A. Frangi. Lecture Notes in Computer Science vol. 9351. Cham, Switzerland: Springer.
- Saretta, E., P. Bonomo and F. Frontini. 2020. A calculation method for the BIPV potential of Swiss façades at LOD2.5 in urban areas: A case from Ticino region. *Solar Energy* 195:150–165.
- Shao, Z., Y. Pan, C. Diao and J. Cai. 2019. Cloud detection in remote sensing images based on multiscale features-convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing* 57 (6):4062–4076.
- Shao, Z., L. Wang, Z. Wang, W. Du and W. Wu. 2020. Saliency-aware convolution neural network for ship detection in surveillance video. *IEEE Transactions on Circuits and Systems for Video Technology* 30 (3):781–794.
- Shao, Z., N. Yang, X. Xiao, L. Zhang and Z. Peng. 2016. A multi-view dense point cloud generation algorithm based on low-altitude remote sensing images. *Remote Sensing* 8 (5):381.
- Shao, Z., L. Zhang and L. Wang. 2017. Stacked sparse autoencoder modeling using the synergy of airborne LiDAR and satellite optical and SAR data to map forest above-ground biomass. *IEEE Journal of Selected Topics in applied Earth Observations and Remote Sensing* 10 (12):5569–5582.
- Shewchuk, J. R. 2002. Delaunay refinement algorithms for triangular mesh generation. *Computational Geometry* 22 (1–3):21–74.
- Simonyan, K. and Z. Zisserman. 2015. Very deep convolutional networks for large-scale image recognition. Pages PP–PP in *3rd International Conference on Learning Representations, ICLR 2015, Conference Track Proceedings*, held in San Diego, Calif., 7–9 May 2015. Edited by J. Editor. City, St.: Publisher.
- Su, H., V. Jampani, D. Sun, S. Maji, E. Kalogerakis, M.-H. Yang and J. Kautz. 2018. SPLATNet: Sparse lattice networks for point cloud processing. Pages 2–4 in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, held in Salt Lake City, Utah, 18–23 June 2018. Edited by J. Editor. City, St.: IEEE.
- Su, H., S. Maji, E. Kalogerakis and E. Learned-Miller. 2015. Multi-view convolutional neural networks for 3D shape recognition. Pages 945–953 in *2015 IEEE International Conference on Computer Vision (ICCV)*, held in Santiago, Chile, 7–13 December 2015. Edited by J. Editor. City, St.: IEEE.
- Teboul, O., I. Kokkinos, L. Simon, P. Koutsourakis and N. Paragios. 2011. Shape grammar parsing via Reinforcement Learning. Pages 2273–2280 in *CVPR 2011*, held in Providence, R.I., 20–25 June 2011. Edited by J. Editor. City, St.: IEEE.
- Wan, G. and A. Sharf. 2012. Grammar-based 3D facade segmentation and reconstruction. *Computers & Graphics* 36 (4):216–223.
- Wang, P., P. Chen, Y. Yuan, D. Liu, Z. Huang, X. Hou and G. Cottrell. 2018. Understanding convolution for semantic segmentation. Pages 1451–1460 in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, held in Lake Tahoe, Nev., 12–15 March 2018. Edited by J. Editor. City, St.: IEEE.
- Wendel, J., A. Simons, A. Nichersu and S. M. Murshed. 2017. Rapid development of semantic 3D city models for urban energy analysis based on free and open data sources and software. Article 15 in *UrbanGIS'17: Proceedings of the 3rd ACM SIGSPATIAL Workshop on Smart Cities and Urban Analytics*, held in Redondo Beach, Calif., November 2017. New York: Association for Computing Machinery.
- Zhang, L., Z. Li, A. Li and F. Liu. 2018. Large-scale urban point cloud labeling and reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing* 138:86–100.
- Zhang, M., L. Zhang, P. Takis Mathiopoulos, Y. Ding and H. Wang. 2013. Perception-based shape retrieval for 3D building models. *ISPRS Journal of Photogrammetry and Remote Sensing* 75:76–91.
- Zheng, Q., A. Sharf, G. Wan, Y. Li, N. J. Mitra, D. Cohen-Or and B. Chen. 2010. Non-local scan consolidation for 3D urban scenes. *ACM Transactions on Graphics* 29 (4):94.

# SUSTAININGMEMBERS

## **Aerial Services, Inc.**

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

## **Ayres Associates**

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

## **Cardinal Systems, LLC**

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

## **Dewberry**

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

## **Esri**

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

## **F.R. Aleman & Associates, Inc.**

Miami, Florida  
https://fr-aleman.com  
Member Since: 7/2020

## **GeoCue Group**

Madison, Alabama  
http://www.geocue.com  
Member Since: 10/2003

## **Geographic Imperatives LLC**

Centennial, Colorado  
http://geographicimperativesllc.com  
Member Since: 9/2021

## **GeoWing Mapping, Inc.**

Richmond, California  
www.geowingmapping.com  
Member Since: 12/2016

## **GPI Geospatial Inc.**

formerly Aerial Cartographics of America, Inc. (ACA)  
Orlando, Florida  
www.aca-net.com  
Member Since: 10/1994

## **Green Grid Inc.**

San Ramon, California  
www.greengridinc.com  
Member Since: 1/2020

## **Half Associates, Inc.**

Richardson, Texas  
www.half.com  
Member Since: 8/2021

## **Keystone Aerial Surveys, Inc.**

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

## **Kucera International**

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

## **L3Harris Corporation**

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

## **Pickett and Associates, Inc.**

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

## **Quantum Spatial, Inc.**

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

## **Sanborn Map Company**

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

## **Surveying And Mapping, LLC (SAM)**

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

## **T3 Global Strategies, Inc.**

Bridgeville, Pennsylvania  
https://t3gs.com/  
Member Since: 6/2020

## **Terra Remote Sensing (USA) Inc.**

Bellevue, Washington  
www.terraremove.com  
Member Since: 11/2016

## **Towill, Inc.**

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

## **Wingtra**

Zurich, Switzerland  
https://wingtra.com/  
Member Since: 6/2020

## **Woolpert LLP**

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

# SUSTAININGMEMBERBENEFITS

## **Membership**

- ✓ Provides a means for dissemination of new information
- ✓ Encourages an exchange of ideas and communication
- ✓ Offers prime exposure for companies

## **Benefits of an ASPRS Membership**

- Complimentary and discounted Employee Membership\*
- E-mail blast to full ASPRS membership per year\*
- Professional Certification Application fee discount for any employee
- Discounts on *PE&RS* Classified Ad
- Member price for ASPRS publications
- Discount on group registration to ASPRS virtual conferences
- Sustaining Member company listing in ASPRS directory/website
- Hot link to company website from Sustaining Member company listing page on ASPRS website
- Press Release Priority Listing in *PE&RS* Industry News
- Priority publishing of Highlight Articles in *PE&RS* plus, 20% discount off cover fee, one per year
- Discount on *PE&RS* full-page advertisement
- Exhibit discounts at ASPRS sponsored conferences (exception ASPRS/ILMF)
- Free training webinar registrations per year\*
- Discount on additional training webinar registrations for employees
- Discount for each new SMC member brought on board (Discount for first year only)

\*quantity depends on membership level

# Scene Classification of Remotely Sensed Images via Densely Connected Convolutional Neural Networks and an Ensemble Classifier

Qimin Cheng, Yuan Xu, Peng Fu, Jinling Li, Wei Wang, and Yingchao Ren

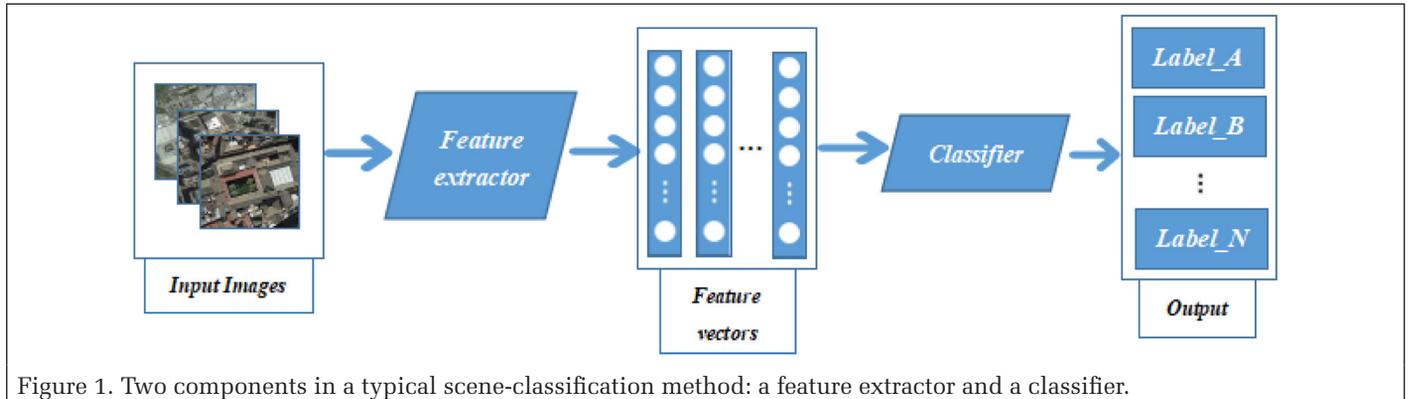


Figure 1. Two components in a typical scene-classification method: a feature extractor and a classifier.

## Abstract

Deep learning techniques, especially convolutional neural networks, have boosted performance in analyzing and understanding remotely sensed images to a great extent. However, existing scene-classification methods generally neglect local and spatial information that is vital to scene classification of remotely sensed images. In this study, a method of scene classification for remotely sensed images based on pretrained densely connected convolutional neural networks combined with an ensemble classifier is proposed to tackle the underutilization of local and spatial information for image classification. Specifically, we first exploit the pretrained DenseNet and fine-tuned it to release its potential in remote-sensing image feature representation. Second, a spatial-pyramid structure and an improved Fisher-vector coding strategy are leveraged to further strengthen representation capability and the robustness of the feature map captured from convolutional layers. Then we integrate an ensemble classifier in our network architecture considering that lower attention to feature descriptors. Extensive experiments are conducted, and the proposed method achieves superior performance on UC Merced, AID, and NWPU-RESISC45 data sets.

## Introduction

Technological advancement has made it possible to acquire high-resolution remotely sensed images at meter and

Qimin Cheng and Yuan Xu are with the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, China (chengqm@hust.edu.cn).

Peng Fu is with the Department of Plant Biology, University of Illinois at Urbana-Champaign, IL 61801.

Jinling Li is with Xunlei Limited, Shenzhen, China.

Wei Wang and Yingchao Ren are with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China; and the College of Resources and Environment, University of Chinese Academy of Sciences, Beijing, China.

submeter levels (such as with SkySat and QuickBird). These remotely sensed high-resolution images contain not only lots of spatial and texture information but also scene semantic information. Due to the enormous number, complex geometric structures, and rich spatial patterns of remotely sensed images, scene classification of these images has become a challenging but active subject in the field of remote sensing (Cheng, Han and Lu 2017). Existing scene-classification methods often consist of two components: a feature extractor and a classifier, as shown in Figure 1. The feature extractor is of importance for scene classification by providing effective image features to a classifier (Lu *et al.* 2017). Early studies in scene classification (Newsam *et al.* 2004; Bhagavathy and Manjunath 2006; Yang and Newsam 2008, 2010; H. Li *et al.* 2010; Cheng, Han, Guo and Liu 2015; Chakraborty *et al.* 2017) were primarily based on handcrafted low-level features (color, texture, edge, etc.), which can seldom satisfy demands for automatic and accurate scene classification in real applications such as natural-hazard monitoring, land cover/land use mapping, geospatial object detection, geographic image retrieval, environment monitoring, and urban planning (Cheng, Han and Lu 2017).

With recent progress in machine learning, many excellent convolutional neural network (CNN) models, such as VGG-Net (Simonyan and Zisserman 2015) and GoogLeNet (Szegedy *et al.* 2015), have been developed and successfully used in computer vision as well as scene classification of remotely sensed images (Heipke and Rottensteiner 2020). The success of CNN-based methods is attributed to their capacity to learn hierarchical representation to describe the image scene. In the past several years, many studies have been performed in CNN-based remote-sensing image scene classification. Cheng, Li *et al.* (2017) developed a novel feature representation for scene

Photogrammetric Engineering & Remote Sensing  
Vol. 87, No. 3, March 2021, pp. 295–308.  
0099-1112/21/295–308

© 2021 American Society for Photogrammetry  
and Remote Sensing  
doi: 10.14358/PERS.87.3.295

classification in remotely sensed images, known as bag of convolutional features, which generated visual words directly from deep convolutional features. Wan *et al.* (2017) employed a local feature-representation scheme based on feature pooling and nonlinear divisive normalization for scene classification. Chaib *et al.* (2017) fused two deep image features to improve the discriminability of semantic objects, which first combined full connected layers of different CNNs to construct the final feature representation of the input image. H. Yu *et al.* (2018) proposed a multi-scale hierarchical structure to represent large-scale images from unmanned aerial vehicles. G. Li *et al.* (2018) proposed a multi-model feature-extraction framework to derive powerful remote-sensing image features for classification. Y. Liu *et al.* (2018) used a new CNN-based scene-classification model to achieve the extraction of high-level and multi-scale features of the objects for high-resolution remote-sensing scene classification. Most recently, a novel feature-representation method introducing a frequency-domain branch to the traditional space-domain-only architecture was proposed by Zheng *et al.* (2019).

Although CNN-based methods have been popular and made remarkable achievements in scene classification of remotely sensed images, limitations still exist. The main challenge is due to the high within-class variation and between-classes ambiguity of remotely sensed data. Most of the previous CNN-based scene-classification methods use outputs from full connection layers of CNNs as the global image features (Castelluccio *et al.* 2015; Luus *et al.* 2015) to represent features conveyed from lower-level layers. However, remotely sensed images with the same global features may have quite dissimilar semantic objects. The use of outputs from fully connected layers as classification features only can lead to scene misclassification. Besides, with the increasing challenges in computer-vision and machine-learning tasks, the models of deep neural networks get more and more complex (Gu *et al.* 2018), which may not satisfy the requirement of real-time applications.

Nogueira *et al.* (2017) explored three strategies to exploit existing CNNs for remote-sensing image scene classification in different scenarios: fully trained CNNs, fine-tuned pretrained CNNs, and pretrained CNNs as feature extractors. Their experiments demonstrated that the fine-tuning strategy tended to perform the best at scene classification. As remotely sensed images contain more semantic objects, CNNs pretrained on natural images to extract deep image features are not tailored to remote-sensing images. In this study, we use a pretrained network—DenseNet (G. Huang *et al.* 2017)—and then fine-tune it using augmented remotely sensed image data to extract remote-sensing image features.

DenseNet, which is a novel deep learning network developed from GoogLeNet Inception (Szegedy *et al.* 2015) and ResNet (K. He *et al.* 2016), uses dense connection and concatenation for feature aggregation to guarantee maximum information communication between layers. Due to its compelling advantages, including reduced gradient vanishing and strengthened feature propagation, DenseNet has since its proposal in 2017 received increasing attention and application in a variety of visual tasks, such as scene classification (K. Zhang *et al.* 2019), superresolution (L. Wang *et al.* 2018; Zhou *et al.* 2018), image caption generation (X. He *et al.* 2019), video action recognition (Hao and Zhang 2019), and human pose estimation (Chu *et al.* 2019). It has especially shown its superiority in the field of medical science (R. Zhang *et al.* 2018; Z. Zhang *et al.* 2018; Bui *et al.* 2019; Dezaki *et al.* 2019; Khened *et al.* 2019; H. Wang *et al.* 2019). Although there exist studies introducing DenseNet into the field of remote sensing for object extraction (X. Li *et al.* 2018), few of them use the network for scene classification. To the best of our knowledge, DenseNet was first used for scene classification of aerial images (Dede *et al.* 2019). In that study, three of ensemble

strategies—snapshot, homogeneous, and heterogeneous—were explored with both DenseNet and Inception networks to show the superiority of ensemble networks over a single network in the context of aerial scene classification. Minetto *et al.* (2019) developed the Hydra framework to automatically create ensembles of CNNs to perform land use classification in satellite images. The Hydra's body was a CNN that was coarsely optimized and then fine-tuned multiple times to form the heads. To prove its validation, ResNet and DenseNet were used to create ensembles. To better fuse the network deepening and widening strategies, C. Zhang *et al.* (2019) utilized the “network in network” concept, in which the idea of DenseNet was introduced to make a deeper network. The validation was tested on 10 kinds of satellite images, but the proposed networks, containing many convolution, pooling, nonlinear, and concatenating layers, were rather complex and time-consuming. Tao *et al.* (2018) proposed a novel depth-width-reinforced deep neural network for per-pixel classification of very-high-resolution remote-sensing images, in which densely connected neural networks and internal classifiers were combined to build a deeper network and balance the network depth and performance. The limitation of that study was the complexity of the network, similar to the work of C. Zhang *et al.* The difference between these DenseNet studies and ours is that we make full use of the advantages of DenseNet to present an effective and efficient feature representation instead of building an ensemble network.

In this study, we exploit a pretrained DenseNet and fine-tune it to release its potential in remote-sensing image scene classification. To further strengthen representation capability and robustness of the feature map captured from convolutional layers, a pyramid structure and a Fisher-vector coding strategy are leveraged. Another important aspect of this study is the consideration of the integration of an ensemble classifier except for feature descriptors. Huge efforts have been made in reinforcing the capability of feature representation via feature descriptors, feature fusion, or feature aggregation. Comparatively less attention has been paid to building ensemble classifiers.

Specifically, our proposed method includes three main parts. First, DenseNet—pretrained by the ImageNet data set—is used to extract visual features. The output of the final dense block is treated as the local feature of the input image. The pretrained DenseNet is further fine-tuned by using augmented remotely sensed images to ensure the robustness of its image-representation capability. Second, deep feature descriptors are trained to obtain a visual dictionary with  $K$  dictionary items. Then, since spatial information is underutilized in existing CNN-based image-classification methods, feature maps extracted from convolutional layers are partitioned into a spatial pyramid with  $m$  blocks in  $N$  pyramid levels. A spatial block is represented as a vector by using the improved Fisher-vector (FV) method. Thus,  $m$  blocks correspond to  $m$  FVs, which are concatenated into a  $2mDk$ -dimensional feature vector as the final image-level representation. Finally, as previous studies usually pay more attention to feature descriptors than classifiers, an ensemble classifier, the AdaBoosting algorithm, is used to train individual base classifiers and stack them all into one by giving them a weight per classification performance. In sum, the main contribution of this work is to improve performance in scene classification of remotely sensed images by providing powerful and discriminative image representation based on a pretrained DenseNet as well as an ensemble classifier. We conducted comparative experiments on three challenging data sets: UC Merced, AID, and NWPU-RESISC45. Experimental results demonstrate the good performance of our classification approach.

The remainder of this article is organized as follows. In the next section we provide a brief overview of previous studies in scene classification of remotely sensed images. After

that, the proposed framework is described in detail. The data analysis and results on several benchmark databases are then presented. Finally, a discussion and conclusions are offered.

## Previous Work

This section comprehensively reviews existing methods for scene classification of remotely sensed images comprehensively. Different from per-pixel or object-based image classification, which explain remotely sensed images in a bottom-up way, scene classification tends to directly model the remote-sensing scene through an overall representation (Cheng *et al.* 2014; Cheng, Han, Guo *et al.* 2015). Scene-classification methods for remote-sensing images can be divided into three types based on the semantic level of features extracted from satellite images: low, mid, and high.

### Scene Classification Based on Low-Level Features

In early years, low-level features such as spectrum, color, texture, structure, and local invariant features were used to classify images. Feature vectors extracted from low-level visual characteristics are often used to describe remotely sensed images either locally (Yang and Newsam 2008) or globally (Aptoula 2014). In real applications, local structure descriptors—e.g., the scale-invariant feature transform (SIFT; Lowe 2004)—are used to describe the complex structure of remote-sensing images. Information from spatial cues such as color and texture (Risojević and Babić 2011) is usually investigated to depict spatial arrangements of remotely sensed images. Dos Santos *et al.* (2010) evaluated various global color descriptors, texture descriptors, and local binary patterns for scene classification. It is worth noting that handcrafted features have limited representation capability and cannot depict highly diverse and nonhomogeneous spatial relationships within remote-sensing scenes well, resulting in high-dimensional and noisy features.

### Scene Classification Based on Mid-Level Features

The limitations of low-level descriptors make it difficult to classify image scenes accurately to meet the needs of practical applications (e.g., natural-hazard monitoring, land cover/land use mapping, geospatial object detection, geographic image retrieval, environment monitoring, and urban planning). Therefore, some previous works (Yang and Newsam 2008; L. Huang *et al.* 2016) attempted to develop a holistic scene representation using high-order statistical patterns formed by the extracted local visual attributes. The main idea underlying mid-level feature-based methods is to encode the local image features—e.g., SIFT and local binary patterns—to build a holistic mid-level representation for remotely sensed images.

Bag of visual words (BoVW) has been considered a mid-level descriptor, which learns image features based on visual dictionaries. It creates a codebook of visual discriminative patches and computes the statistical word frequency of the test image. Because of its simplicity and efficiency, the BoVW descriptor and its variants have been state-of-the-art for many years in computer-vision tasks (Csurka *et al.* 2004; Faraji and Shanbehzadeh 2015) and have been widely adopted to compute mid-level representations for remotely sensed images (Negrel *et al.* 2014; L. Zhao *et al.* 2014a, 2014b). There have been many other methods used to encode low-level features to mid-level features, such as FV (Sánchez *et al.* 2013), vector of locally aggregated descriptors (Jégou *et al.* 2010), and sparse coding (Cheriyadat 2013; Taigman *et al.* 2014).

To encode high-order spatial information from low-level local visual words for scene modeling, topic models based on the BoVW scheme are developed to model the semantic relationship among the visual words (Lienou *et al.* 2010; Hu *et al.* 2013; Kusumaningrum *et al.* 2014; Zhong *et al.* 2015). These

models include the gray-level co-occurrence matrix (Haralick *et al.* 1973), CIE Lab color moments (Stricker and Orengo 1995), latent Dirichlet allocation (Blei *et al.* 2003), edge-orientation histogram, probabilistic latent semantic analysis (Bosch *et al.* 2006), spatial-pyramid matching kernel (Lazebnik *et al.* 2006), and spatial-pyramid coexistence kernel (Yang and Newsam 2011). However, all the mid-level features built for scene classification are based on handcrafted local image descriptors which cannot fully describe complex spatial structures of images. Plus, feature-extraction algorithms rely greatly on expert knowledge for the design of handcrafted local image features, rendering them inflexible and inadaptable to scenes in different geographic regions. Therefore, methods based on semi-supervised or unsupervised feature learning are alternatives.

### Scene Classification Based on High-Level Features

In recent years, a variety of CNN-based methods (Castelluccio *et al.* 2015; Luus *et al.* 2015; Chaib *et al.* 2017; Cheng, Li *et al.* 2017; G. Huang *et al.* 2017; Wan *et al.* 2017; G. Li *et al.* 2018) have been popular in the field of remote-sensing image scene classification (Cheng *et al.* 2018). Deep learning-based classification methods can learn more abstract and discriminative semantic features and yield better classification performance (Castelluccio *et al.* 2015; Luus *et al.* 2015; Hu *et al.* 2015; F. Zhang *et al.* 2016) than methods based on low-level and mid-level features. This better performance should be attributed to the fact that deep learning-based methods use a multi-stage global feature-learning architecture to adaptively learn image features and often cast scene classification as an end-to-end problem. Deep learning-based scene-classification methods for remotely sensed images can be further divided into two categories: methods based on pretrained CNNs and on re-trained CNNs.

#### Methods Based on Pretrained CNNs

Using deep neural networks pretrained on natural images (Russakovsky *et al.* 2015), Penatti *et al.* (2015) showed that deep features of daily objects not only can be generalized to remote-sensing domains but also exhibit impressive performance on remote-sensing scene classification. Nogueira *et al.* (2017) evaluated six pretrained deep CNN architectures—AlexNet (Krizhevsky *et al.* 2012), OverFeat (Sermanet *et al.* 2014), CaffeNet (Jia *et al.* 2014), VGG16 (Simonyan and Zisserman 2015), and GoogLeNet (Szegedy *et al.* 2015)—for scene classification of remotely sensed images. Alternatively, a pretrained network can be used as a feature extractor. Studies have shown that CNN-based feature extractors can be successfully translated to computer-vision tasks (Girshick *et al.* 2014; W. Yu *et al.* 2014). When a pretrained CNN is applied to image classification, feature maps in higher layers associated with semantic information are selected to represent the whole image. Scott *et al.* (2017) used a multi-layer perceptron as a classifier with input features extracted from a fully connected layer of pretrained CNNs. Marmanis *et al.* (2016) composed features extracted from a fully connected layer into a 2D array and then fed them into a neural network to obtain the classification results. B. Zhao *et al.* (2017) trained a multi-layer perceptron as a classifier with input features extracted from pretrained networks and then connected it with the feature-extraction part of the pretrained networks. Their experimental results showed that classification performance improved slightly, but an additional data set was needed to build a fully pretrained network. Hu *et al.* (2015) extracted features from the fully connected layers from different CNNs and concentrated these features into a single feature vector. Sheng *et al.* (2012) presented a new classification approach based on the sparse codes of multiple features for high-resolution satellite scenes. Their approach generalized vector quantization to sparse coding and used a two-stage linear

support vector machine classifier to effectively compose a set of diverse and complementary features.

In addition to features of fully connected layers, convolutional-layer features have attracted attention. J. Wang *et al.* (2017) designed a linear principal-component analysis (PCA) network to synthesize spatial information of remotely sensed images in each spectral channel before feeding them to pretrained deep CNNs. Cheng *et al.* (2016) and Huo *et al.* (2017) made efforts to address rotation and scale variations in remote-sensing images. Mahendran and Vedaldi (2015) and Yosinski *et al.* (2015) proved that outputs from layers other than fully connected ones of CNNs could be used as image features for scene classification. Hu *et al.* (2015) used multi-scale dense CNN activations from the last convolutional layer as local feature descriptors, which were further coded using feature-encoding methods such as BoVW, vector of locally aggregated descriptors, and improved Fisher kernel to generate the final image representation. Q. Liu *et al.* (2018) proposed a multi-scale deep feature-learning method for high-resolution satellite-image scene classification, in which pretrained spatial-pyramid pooling was employed to automatically generate a multi-scale fixed-length feature representation regardless of image size or scale. W. Zhang *et al.* (2019) proposed an effective remote-sensing image scene-classification architecture to make full use of the strengths of both CNNs and CapsNet, using CNNs without fully connected layers as an initial feature map extractor. Flores *et al.* (2019) utilized a pretrained deep CNN and a Gaussian mixture model to generate more robust and compact dictionaries of deep features.

#### Methods Based on Retrained CNNs

In addition to scene classification based on pretrained CNNs, CNNs can be trained from scratch or fine-tuned from pretrained networks. To better adapt to small remote-sensing image data sets, small-size classification networks have been trained (Yang and Newsam 2011). F. Zhang *et al.* (2016) developed a gradient-boosting random convolutional network for

classifying remotely sensed images with only two convolutional layers. Marmanis *et al.* (2016) observed improved classification accuracy using fine-tuned pretrained networks; they compared CaffeNet and GoogLeNet for scene classification (using the UC Merced data set) in three different ways—re-training from scratch, fine-tuning, and using pretrained CNNs as a feature extractor.

### The Proposed Method

The overall flowchart of the proposed method is shown in Figure 2. First, DenseNet is used to extract image features. The output of the last dense block in DenseNet is used as the local feature. Second, a visual dictionary is generated by training a Gaussian mixture model (GMM) with local features extracted from training images. Then a spatial-pyramid structure is used to partition the feature map into  $m$  blocks in  $N$  pyramid levels. Each block feature map is encoded into a single vector using the improved FV. After that, the  $m$  single vectors are concatenated to a  $2mDk$ -dimensional feature vector as the final image representation. Finally, an ensemble classifier is built to perform classification.

#### Feature Extraction

DenseNet (G. Huang *et al.* 2017), developed from ResNet and the Inception structure of GoogLeNet, essentially has a brand-new network structure that is effective for scene classification. In recent years, there have been two main directions for optimizing CNN performance in tasks such as image classification and pattern recognition: deepening or widening the network structure. In CNNs, as the network depth increases, gradient disappearance becomes obvious. New networks have been devised to solve this problem, such as ResNet, highway networks, stochastic depth, and FractalNet. Although the structures of these networks are different, the core idea is the same: creating short paths from layer to layer. In DenseNet this idea is continued, as the input of each layer comes from

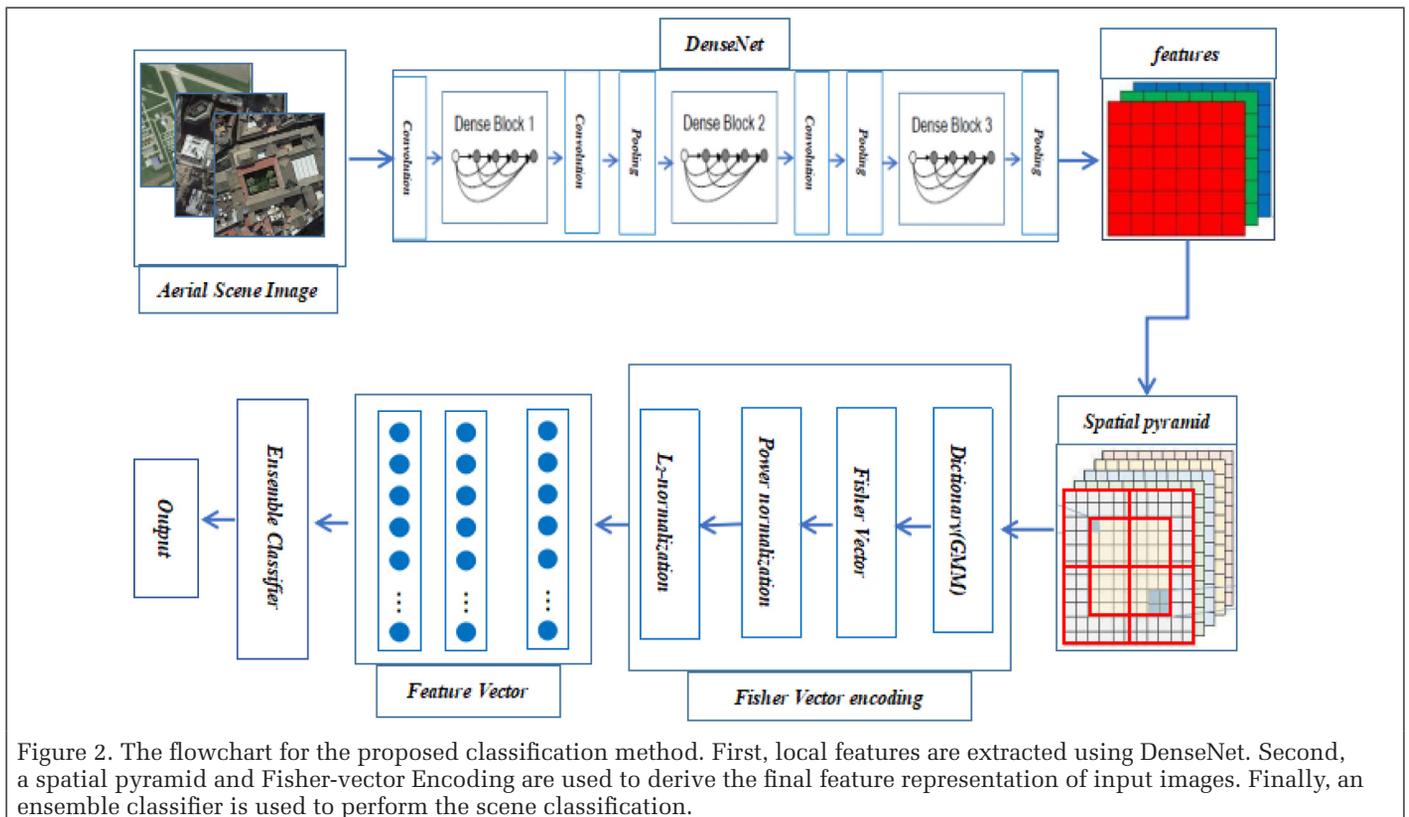


Figure 2. The flowchart for the proposed classification method. First, local features are extracted using DenseNet. Second, a spatial pyramid and Fisher-vector Encoding are used to derive the final feature representation of input images. Finally, an ensemble classifier is used to perform the scene classification.

the synthesized outputs of all previous layers. With this kind of dense structure, all layers in DenseNet are connected, which ensures maximum information transmission between layers in the network.

Consider a single image  $X_0$  that is passed through a convolutional network. The network comprises  $L$  layers, each of which implements a nonlinear transformation  $H_l(\cdot)$  for layer  $l$ .  $H_l(\cdot)$  can represent a composite function of operations such as batch normalization, rectified linear units, pooling, or convolution.  $X_l$  denotes the output of layer  $l$ . Traditional convolutional feed-forward networks connect the output of layer  $l$  as input to layer  $l + 1$ , which gives rise to the following layer transition:  $X_l = H_l(X_{l-1})$ . Compared to traditional convolutional feed-forward networks, ResNet adds a skip connection that bypasses the nonlinear transformations with an identity function:

$$X_l = H_l(X_{l-1}) + X_{l-1}. \quad (1)$$

As a result, the advantage of ResNet is that the gradient can flow directly through the identity function from back to front layers. However, the identity function and the output of  $H_l$  are combined by summation, which may impede the information flow in the network. To further improve the information flow between layers, DenseNet uses a different connectivity pattern. The structure of DenseNet is shown in Figure 3. Layer  $l$  receives the feature maps of all preceding layers,  $X_0, \dots, X_{l-1}$ , as the input of the current layer:

$$X_l = H_l([X_0, X_1, \dots, X_{l-1}]), \quad (2)$$

where  $[X_0, X_1, \dots, X_{l-1}]$  refers to the concatenation of the feature maps produced in layers  $0, \dots, l-1$ . Due to the dense block design, DenseNet has a narrow network structure, with only a few parameters. The number of output feature maps per convolution layer in the dense block is less than 100, unlike the thousands of feature maps in other networks. At the same time, the dense connection makes it more efficient for the transmission of features and gradients. In this study, we

chose DenseNet as the feature extractor. As feature maps extracted by DenseNet have information from each layer, these features are expected to better describe the training images than those from networks in which the input of each layer is related only to the output of a previous layer. DenseNet was trained using the ImageNet data set with four dense blocks. Given the differences between remotely sensed images and those in the ImageNet data set, the pretrained DenseNet was further fine-tuned for feature extraction.

### Feature Processing

The convolutional features extracted from DenseNet cannot be directly used by the classifier for scene classification, because convolutional-layer features are of different sizes. After extraction of feature maps, these features are normalized to remove outlier features for classification. A spatial pyramid is used to partition features, which are later aggregated into a single vector using FV. Finally, the single feature vectors from all feature blocks are used as the final image representation.

### Feature Normalization

Let  $X = [x_1, \dots, x_i, \dots, x_T]^T$  ( $X_n \in R^{T \times d}$ ) be the image features extracted from image  $I$  by DenseNet. These image features are usually transformed by PCA to improve the overall classification performance (e.g., PCA-transformed SIFT features). However, in practical applications, PCA-transformed features often lose information. In this article, we use two normalization methods,  $\ell_2$  vector normalization (Equation 10) and  $\ell_2$  matrix normalization (Gao *et al.* 2015), using the following equations:

$$x_i = x_i / |x_i|, \quad (3)$$

$$x_i = x_i / \|X\|_2. \quad (4)$$

In Equation 3,  $|x_i|$  represents the modulus of the vector  $x_i$ , and in Equation 4,  $\|X\|_2$  represents the spectral norm of the matrix  $X$ —i.e., the maximum singular values of the matrix  $X$ . Compared to  $\ell_2$  vector normalization,  $\ell_2$  matrix normalization is the normalization of features by the information of the entire image.

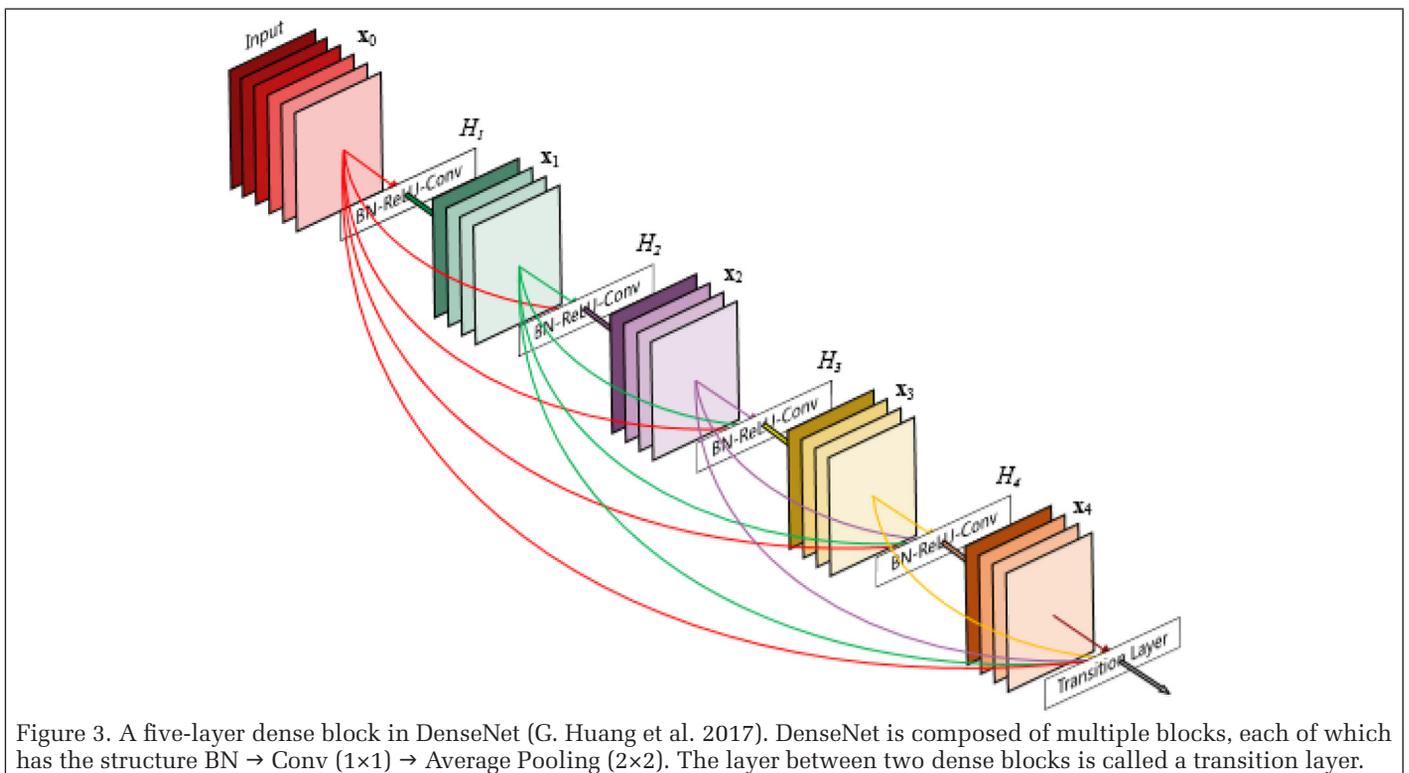


Figure 3. A five-layer dense block in DenseNet (G. Huang et al. 2017). DenseNet is composed of multiple blocks, each of which has the structure BN → Conv (1×1) → Average Pooling (2×2). The layer between two dense blocks is called a transition layer.

### Image Representation Based on a Spatial Pyramid

The main function of a spatial pyramid is to capture the spatial patterns in an image. It has been shown that the synergistic use of spatial and spectral information can significantly improve image-classification performance (Lazebnik *et al.* 2006). He *et al.* (2015) showed that a sparse spatial-pyramid network improved recognition performance by adding a spatial-pyramid pooling layer. A more intuitive and natural idea that they proposed is to aggregate features extracted from a CNN into an FV. Thus, in this study, we use a spatial pyramid to capture the spatial information in an image and then use an FV to aggregate the features in each block of the spatial pyramid into a single feature vector (Table 1).

As mentioned before, feature maps derived from neural networks can be used as local image features. In a feature map, a single cell corresponds to a local image patch in the input image, and the arrangement of all feature-map cells form a regular grid of image patches in the input image. Thus, it is easy to form a spatial pyramid by partitioning an image into multiple subregions and then calculating local features in each subregion (Figure 4). The level 0 spatial pyramid aggregates all the local features. The level 1 spatial pyramid, as shown in the red part of Figure 4, divides the entire feature map into five subregions: four quadrant regions and one central region. The features in each subregion are then aggregated by FV; eventually, five aggregated single vectors will be generated. In practical applications, the number of layers of the spatial pyramid is determined by the task and an expert. In this study, given the dimension of the final image representation and the computational cost, only two spatial-pyramid layers are used: levels 0 and 1. Therefore, a total of six single vectors are finally generated to form the final representation for an image.

### Fisher Vector

Unlike BoF and vector of locally aggregated descriptors, which use  $K$ -means to cluster image features, the FV method constructs a visual dictionary with a GMM. Although the GMM method is also a clustering method, it considers the distance from a feature point to each cluster center—i.e., the linear combination of all cluster centers is used for image representation. The generic framework of a Fisher kernel combines the benefits of generative and discriminative approaches (Perronnin *et al.* 2010). According to the framework, a matrix of like images can be further represented by a single vector through FV. Let  $I_i$  denote an input scene image; then the set of convolutional features of  $I_i$  can be represented by  $X = \{I_i\} = \{x_1, x_2, \dots, x_T\} \in R^D$ . Let  $\lambda = \{w_i, \mu_i, \Sigma_i, i = 1, 2, \dots, k\}$  be the parameters of a GMM fitting the distribution of features from the convolutional-feature set  $\{I_i\}$ , where  $w_i$ ,  $\mu_i$ , and  $\Sigma_i$  are, respectively, the mixture weight, mean vector, and covariance matrix of the Gaussian model;  $k$  is the component number of the GMM; and  $M$  is the number of scene images. It is assumed that the generation process of  $X$  can be modeled by a probability density function  $P(x|\lambda)$  with the parameters  $\lambda$ . In this case,  $X$  can be described by the gradient vector

$$G_\lambda^x = \frac{1}{T} \sum_{t=1}^T \nabla_\lambda \log p(x_t|\lambda), \quad (5)$$

where

$$p(x_t|\lambda) = \sum_{i=1}^k w_i p_i(x_t|\lambda), \quad (6)$$

$$p_i(x_t|\lambda) = \frac{\exp\left\{\frac{1}{2}(x_t - \mu_i)^T \Sigma_i^{-1} (x_t - \mu_i)\right\}}{(2\pi)^{D/2} |\Sigma_i|^{1/2}}. \quad (7)$$

Table 1. Steps for extracting image features using a spatial-pyramid structure.

Algorithm 1: Build a spatial pyramid

**Input:** The feature map  $X = [x_1, \dots, x_n, \dots, x_N]^T$  of image  $I$ .

**Output:** The final image representation  $f(X)$ .

**Procedure:**

- 1) For each activation vector  $x_n$ , perform  $\ell_2$  matrix normalization  $x_n \leftarrow x_n / X_2$ .
- 2) Use the training set to estimate a Gaussian mixed model  $\lambda = \{\omega_k, \mu_k, \sigma_k\}$ .
- 3) Generate a spatial pyramid  $\{X_1, \dots, X_m\}$  for  $X$ .
- 4) For each  $X_i (1 \leq i \leq m)$ , the Fisher vector is used to generate the corresponding feature vector  $f(X_i)$ .
- 5) Concatenate  $f(X_i) (1 \leq i \leq m)$  to form the final spatial-pyramid representation  $f(X)$ .

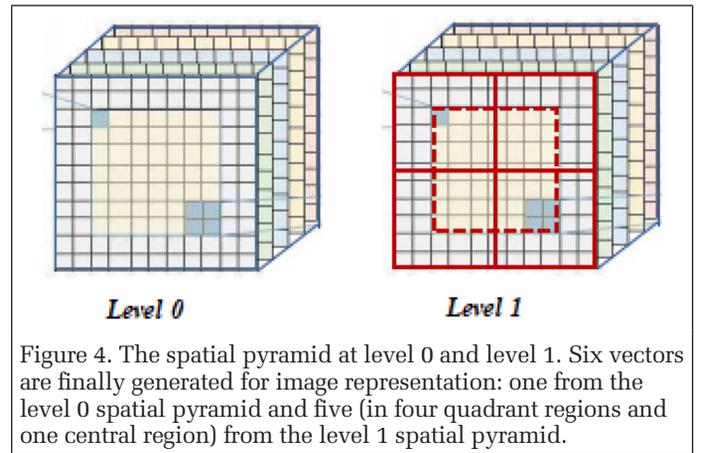


Figure 4. The spatial pyramid at level 0 and level 1. Six vectors are finally generated for image representation: one from the level 0 spatial pyramid and five (in four quadrant regions and one central region) from the level 1 spatial pyramid.

Accordingly, the Fisher vector  $g_\lambda^x$  of  $X$  is defined as

$$g_\lambda^x = F_\lambda^{-1/2} G_\lambda^x, \quad (8)$$

where  $F_\lambda$  is the Fisher information matrix of the GMM model  $\lambda$ . According to Perronnin *et al.* (2010), the  $D$ -dimensional gradients with respect to the mean  $\mu_i$  and standard deviation  $\sigma_i$  are defined as following

$$g_{\mu,i}^x = \frac{1}{T \sqrt{\omega_i}} \sum_{t=1}^T \gamma_t(i) \left( \frac{x_t - \mu_i}{\sigma_i} \right), \quad (9)$$

$$g_{\sigma,i}^x = \frac{1}{T \sqrt{2\omega_i}} \sum_{t=1}^T \gamma_t(i) \left[ \left( \frac{x_t - \mu_i}{\sigma_i} \right)^2 - 1 \right], \quad (10)$$

$$\gamma_t(i) = \frac{\omega_i p_i(x_t|\lambda)}{\sum_{j=1}^k \omega_j p_j(x_t|\lambda)}. \quad (11)$$

In Equations 9–11,  $\gamma_t(i)$  is the occupancy probability of  $x_t$  to the  $i$ th Gaussian component,  $\sigma_i$  is the standard deviation, and  $\sigma_i^2 = \text{diag}(\Sigma_i)$ . The FV theory explains three gradients; here only two—with respect to the mean and standard deviation—are used to construct the FV representation, since the gradient with respect to the mixture weight  $\omega_i$  has little effect on image-recognition results and is always discarded in practical applications. As a result, an FV can be denoted as  $\Phi(X) = \{g_{\mu,1}^x, g_{\sigma,1}^x, \dots, g_{\mu,k}^x, g_{\sigma,k}^x\}^T$ , where  $k$  is the number of GMM components. The final FV of an image is the concatenation of the gradients  $g_{\mu,k}^x$  and  $g_{\sigma,k}^x$  for all  $k$  Gaussian components. Thus, an image can be represented by a set of deep descriptors  $X$  with a

2Dk-dimension FV. In addition, the FV is improved by power normalization with the factor of 0.5, followed by  $L_2$  vector normalization.

### An Ensemble Classifier

An ensemble classifier is a high-performance classifier created by appropriately stacking a plurality of classifiers. In this study we build an ensemble classifier to improve the classification accuracy. Figure 5 outlines the three steps used to build an ensemble classifier. First, base classifiers are individually trained with the training data set. Then each base classifier makes its own predictions for a test data set. Finally, predictions from all base classifiers are stacked to yield final classification results.

This study constructs a simple weighted voting classifier by using the AdaBoosting algorithm. The AdaBoost.M1 algorithm has been favored in some classification competitions in recent years. The specific steps of the algorithm are shown in Table 2. It gives each training sample a weight representing the influence of the sample on the final classification model. The sample weight also determines the probability that a sample will be selected as a training sample. Before training, the weight of each sample is the same. According to the sample weight, sample subsets are obtained for training the base classifiers and then evaluating them. During each epoch of training, the sample weights are adjusted according to the learning results of the base classifier—that is, the weights of misclassified samples are unchanged, whereas those of correctly classified ones decrease. In this way, when the training samples are chosen in the next training epoch, the misclassified samples will receive more attention. By repeating this process, base classifiers are trained. Finally, when the ensemble classifier is applied to the test samples, the weights of each base classifier are combined to predict the categorical results.

## Experiments

### Experimental Data Sets

Three public high-spatial-resolution image data sets are used for evaluating the proposed method. The three data sets are described here and further summarized in Table 3.

#### UC Merced

The UC Merced land use data set (Yang and Newsam 2010) contains 2100 overhead scene images divided into 21 scene-level land use classes. Each class consists of 100 aerial images of 256×256 pixels with a spatial resolution of 0.3 m. This data set is manually extracted from a large number of orthoimages downloaded from the United States Geological Survey

Table 2. Description of the AdaBoost.M1 algorithm.

Algorithm 2: AdaBoost.M1

**Input:** D = data set  $\{x_i, y_i\}$  ( $i = 1, \dots, N$ ).

I = a series of classifiers.

M = number of iterations.

**Output:** The ensemble classifier  $G(x)$ .

**Procedure:**

1) Initialize the weights of training data. The initial weight of each sample is  $w = 1/N$ , where N is the number of samples. We have

$$D_1 = (w_{11}, w_{12}, \dots, w_{1i}, \dots, w_{1N}), w_{1i} = \frac{1}{N} \quad (i = 1, 2, \dots, N), \text{ where } D_1 \text{ represents sample weight}$$

vectors at the first iteration and  $w_{1i}$  represents the weight of the  $i$ th sample at the first iteration.

2) Make multiple iterations  $m = 1, 2, \dots, M$ , where  $m$  represents the number of iterations. In each iteration:

a) Obtain the training set  $D_m$  according to the sample weights.

b) Train the base classifier  $g_m$  with  $D_m$ .

c) Calculate the classification error of the base classifier  $g_m$ —i.e.  $\text{error}(g_m)$ .

d) If  $\text{error}(g_m)$  is greater than 0.5, then  $M = m - 1$ , and the iterative process finishes. Otherwise, the training data weights are updated per the following equations for the next iteration:

$$a_m = \frac{1}{2} \ln \frac{1 - \text{error}(g_m)}{\text{error}(g_m)},$$

$$D_{m+1} = (w_{m+1,1}, w_{m+1,2}, \dots, w_{m+1,i}, \dots, w_{m+1,N}),$$

$$w_{m+1,i} = \frac{w_m}{Z_m} \exp(-a_m y_i g_m(x_i)) \quad (i = 1, 2, \dots, N),$$

$$Z_m = \sum_{j=1}^N w_{mj} \exp(-a_m y_j g_m(x_j)),$$

where  $a_m$  is the base classifier weight,  $D_{m+1}$  is the weight of the sample for the next iteration,  $w_{m+1,i}$  is the weight of the  $i$ th sample at the next iteration,  $y_i$  represents the category corresponding to the  $i$ th sample,  $G_m(x_i)$  represents the classification of the sample  $x_i$  by the base classifier, and  $Z_m$  is a normalization factor such that the sum of the weights of all samples is 1.

3) After completion of iterations, all the base classifiers are stacked to build an ensemble classifier:

$$G(x) = \text{sign} \left( \sum_{m=1}^M a_m g_m(x) \right).$$

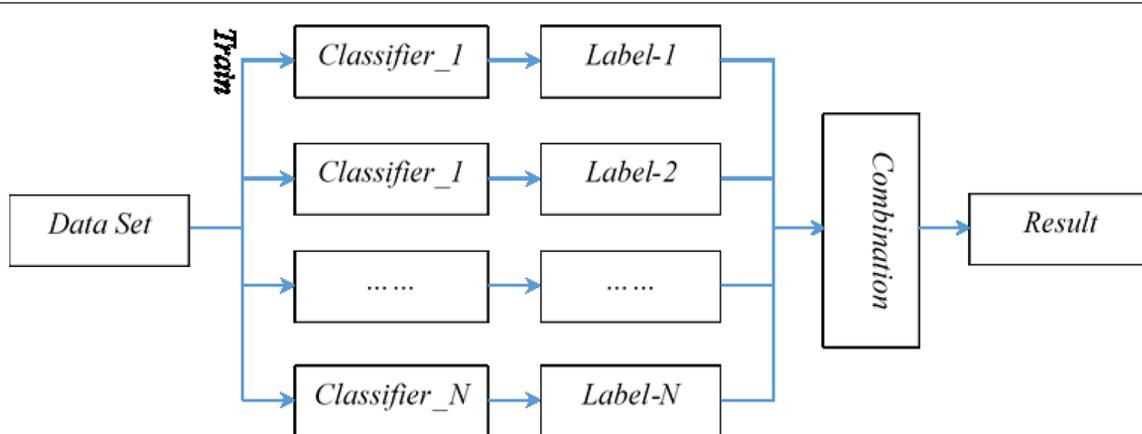


Figure 5. The three steps used to perform ensemble classification.

Table 3. Summary of three data sets used for benchmarking the developed scene-classification method.

Data Set	Images/Class	Scenes	Total Images	Spatial Resolution (m)	Size
UC Merced (Yang and Newsam 2010)	100	21	2100	0.3	256×256
AID (Xia <i>et al.</i> 2017)	220–240	30	10 000	0.5–8	600×600
NWPU-ESISC45 (Cheng, Han and Lu 2017)	700	45	31 500	0.2–30	256×256

National Map for 20 US cities. It has been widely used for evaluating the performance of methods in scene classification of remotely sensed images.

#### AID

The AID data set (Xia *et al.* 2017) is collected from Google Earth and made up of 30 aerial scene categories. The sample number of each class varies from 220 to 420, and the data set contains 10 000 images in total. The pixel resolution ranges from 8 to 0.5 m, and the spatial size of an image is 600×600 pixels.

#### NWPU-RESISC45

The NWPU-RESISC45 data set (Cheng, Han and Lu 2017) consists of 31 500 remote-sensing images divided into 45 scene classes. Each class contains 700 images with a size of 256×256 pixels. As with the AID data set, this data set is also extracted from Google Earth. The 31 500 images cover more than 100 countries and regions all over the world. Compared with other existing scene-classification data sets, this set contains much more data, richer image variation, and higher intraclass diversity and interclass similarities. This data set contains not only changes in viewpoint, translation, object pose and appearance, spatial resolution, illumination, background, and occlusion for each category but also some fine-grained categories with high semantic overlapping, such as circular and rectangular farmlands, commercial and industrial areas, and basketball and tennis courts.

#### Computing Environment and Model Configuration

Data processing and analysis in this study were completed using the 64-bit Ubuntu operating system with 32 GB running memory, a 2-TB hard disk, a 12-GB Titan Xp graphics card, and Python 3.5. The evaluation metric for the proposed method was the overall accuracy—i.e., the number of images correctly classified as a percentage of the total number of test images. The proposed method for scene classification was divided into three parts: fine-tuning of the pretrained DenseNet, image feature extraction, and classifier construction.

#### Fine-Tuning of DenseNet

Before feature extraction, the pretrained DenseNet using the ImageNet images was fine-tuned to adapt to remotely sensed images. The pretrained DenseNet selected in our study had four dense blocks (i.e., DenseNet201; G. Huang *et al.* 2017), with network parameters in each layer shown in Table 4. The fine-tuning began with initial parameter values of the pretrained DenseNet201 and was achieved in subsequent 15-epoch training using images in the training sets. The batch size was set to 128. In addition, to enhance the robustness of the network, various transformations were applied to the original images, including image transpose and rotations. Examples of data enhancement are shown in Figure 6, where Figure 6a shows the original image, Figure 6b and 6c are the images flipped horizontally and vertically, and Figure 6d–6f are images rotated at 90°, 180°, and 270°. In this way, the number of the original image increased by five times. Labels and the associated seed images were linked to the new augmented data sets, allowing a massive expansion of the available labeled training data. Figure 7 shows cross-entropy loss (Figure 7a) and overall accuracy (Figure 7b) of the training process (blue curve) and the validation process (red curve). With longer training (increasing epochs), the performance improves, and the loss and accuracy converge after 15 epochs of training.

Table 4. Specifications of network parameters in DenseNet.

Layer	DenseNet201	Output Size
Convolution	7×7 conv, stride 2	128×128
Pooling	3×3 max pool, stride 2	64×64
Dense block (1)	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	64×64
Transition layer (1)	1×1 conv 2×2 average pool, stride 2	64×64 32×32
Dense block (2)	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	32×32
Transition layer (2)	1×1 conv 2×2 average pool, stride 2	32×32 16×16
Dense block (3)	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$	16×16
Transition layer (3)	1×1 conv 2×2 average pool, stride 2	16×16 8×8
Dense block (4)	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	8×8
Classification	7×7 global average pool 1000D fully connected, softmax	1×1

#### Feature Extraction and Processing

Features used for scene classification in this study were outputs of the connection layer of the last dense block in DenseNet201. As the original image size was 256×256, the corresponding size of convolution feature maps was 8×8×1920.

Before the spatial-pyramid construction, extracted deep features were normalized to eliminate the influence of singular features. The spatial-pyramid structure was used to capture the spatial pattern of an image: the image feature map was divided into six subregions with a two-layer spatial pyramid. The FV was used to compose local features from each subarea. As the number of the local features used to train the GMM was  $8 \times 8 = 64$  (less than 100), the  $K$  value of the GMM was set to 2. Since the AdaBoosting algorithm assumes that base classifiers support the sample weighting, the three classifiers of BP network, KNN, and Bayes were selected.

## Results

This section presents the accuracy assessment of a series of scene-classification methods using the UC Merced and AID data sets. Three additional experiments using the DenseNet201 were conducted, based on the pretrained DenseNet201, the fine-tuned DenseNet201, and the proposed method.

Table 5 shows that DenseNet is superior to VGG, GoogLeNet, ResNet, and other networks in scene classification of remotely sensed images. Compared with the performance of the pretrained DenseNet201, the proposed method improved classification accuracy by 3.41% and 2.02% on the UC Merced and the AID data sets, respectively. Compared with

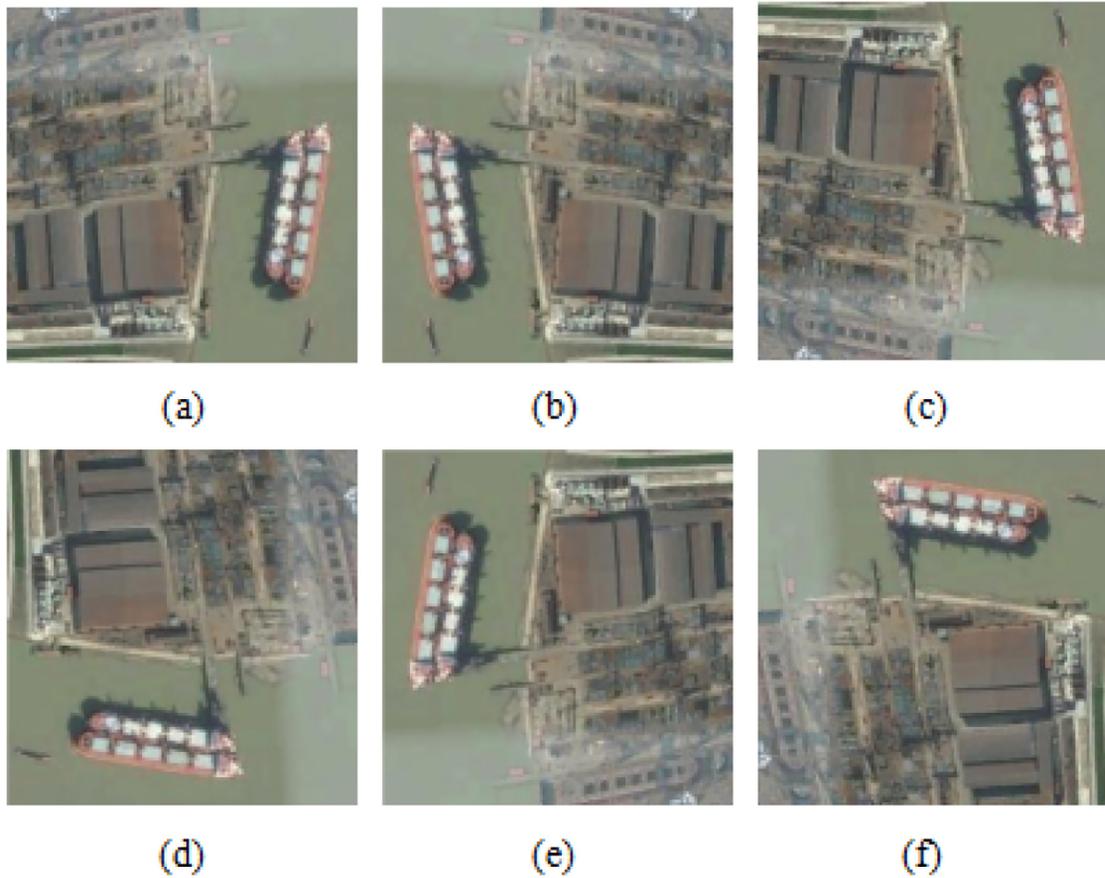


Figure 6. An image in the NWPU-RESISC45 data set with a few rotations: (a) the original image, (b) the original image flipped horizontally, (c) the original image flipped vertically, (d) the original image rotated 90°, (e) the original image rotated 180°, and (f) the original image rotated 270°.

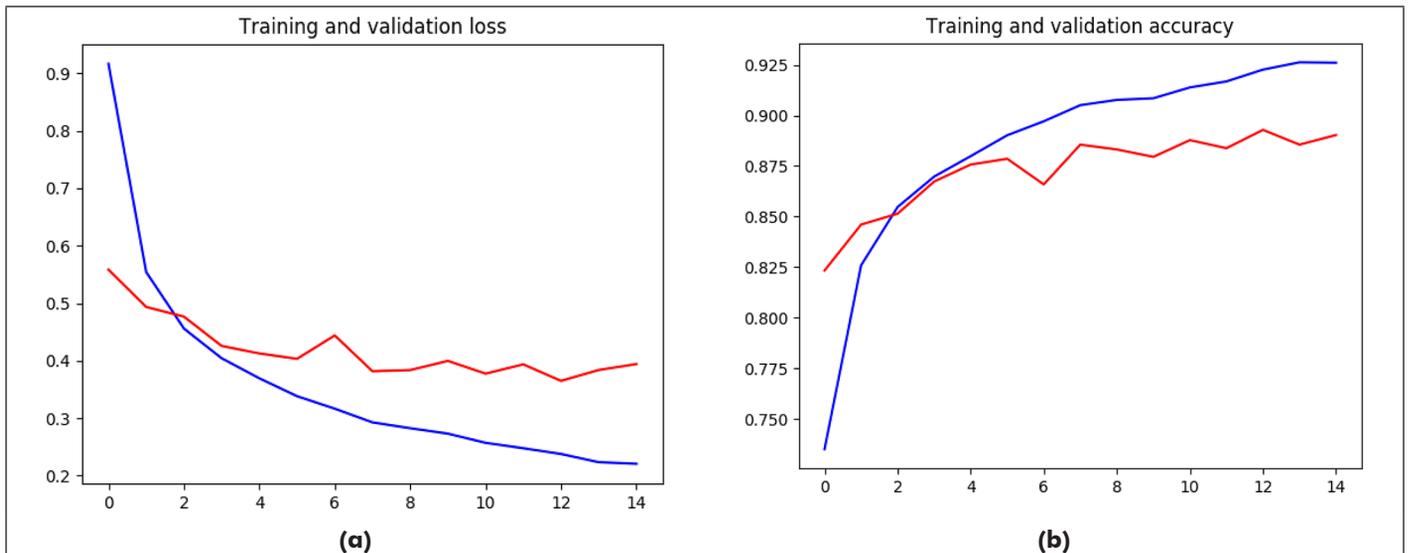


Figure 7. Training results for DenseNet: (a) loss of the training and validation set; (b) accuracy of the training and validation set.

the performance of the pretrained DenseNet201 with fine-tuning and data augmentation, the proposed method improved classification accuracy by 1.8% and 1.22%, respectively. The proposed method yielded the highest classification accuracy: 98.76% and 95.83% on the UC Merced and AID data sets, respectively (Figures 8 and 9).

Comparisons of the proposed method and existing methods for scene classification were also performed based on the NWPU-RESISC45 data set. Table 6 shows that the proposed method achieves the highest scene-classification accuracy compared to existing classification methods on the NWPU-RESISC45 data set. The classification accuracy for each class of the

Table 5. Accuracy assessment of scene-classification methods using the UC Merced and AID data sets.

Method	UC Merced		AID		Notes
	OA (%)	SD	OA (%)	SD	
SIFT	42.10	0.95	26.67	0.56	
LBP	46.29	0.90	36.62	0.49	
Pretrained OverFeatL & Caffe + feature fusion (Penatti <i>et al.</i> 2015)	97.43	0.27	—	—	
Pretrained OverFeat + feature fusion (Marmanis <i>et al.</i> 2016)	92.4	—	—	—	Spatial arrangement
Pretrained CaffeNet (Marmanis <i>et al.</i> 2016)	95.02	0.81	89.53	0.31	Data augmentation
Pretrained VGG16	95.21	1.20	89.64	0.36	
Pretrained AlexNet	95.00	0.52	90.11	0.45	
Pretrained GoogleNet	94.31	0.89	86.39	0.55	
Pretrained AlexNet + fine-tuning (B. Zhao <i>et al.</i> 2017)	96.67	—	—	—	Data augmentation
Pretrained VGGNet + fine-tuning	98.10	0.33	93.67	0.62	
Pretrained ResNet50 + fine-tuning (Scott <i>et al.</i> 2017)	98.50	—	—	—	Data augmentation
Pretrained GoogleNet + fine-tuning (Castelluccio <i>et al.</i> 2015)	97.10	—	—	—	Data augmentation
Pretrained GoogleNet + training from scratch (Castelluccio <i>et al.</i> 2015)	92.86	—	—	—	Data augmentation
Liner PCA + pretrained GoogleNet (Hu <i>et al.</i> 2015)	95.94	0.59	—	—	
Liner quaternion PCA + pretrained GoogleNet (J. Wang <i>et al.</i> 2017)	95.76	—	—	—	
Pretrained VGG16 + 1st-FC (Hu <i>et al.</i> 2015)	96.88	—	—	—	Data augmentation
Pretrained PlacesNet + 1st-FC (Hu <i>et al.</i> 2015)	94.90	—	—	—	Data augmentation
Pretrained VGG16 + feature coding (BOW) (Hu <i>et al.</i> 2015)	96.51	—	—	—	
Pretrained VGG16 + feature coding (LLC) (Hu <i>et al.</i> 2015)	95.64	—	—	—	
Pretrained DenseNet201	95.35	0.41	93.81	0.56	
Pretrained DenseNet201 + fine-tuning	96.96	0.54	94.61	0.23	Data augmentation
Proposed method	98.76	0.16	95.83	0.41	Data augmentation

BOW = XXXX; LBP = XXXX; LLC = XXXX; OA = overall accuracy; PCA = principal-component analysis ; SD = standard deviation; SIFT = scale-invariant feature transform.

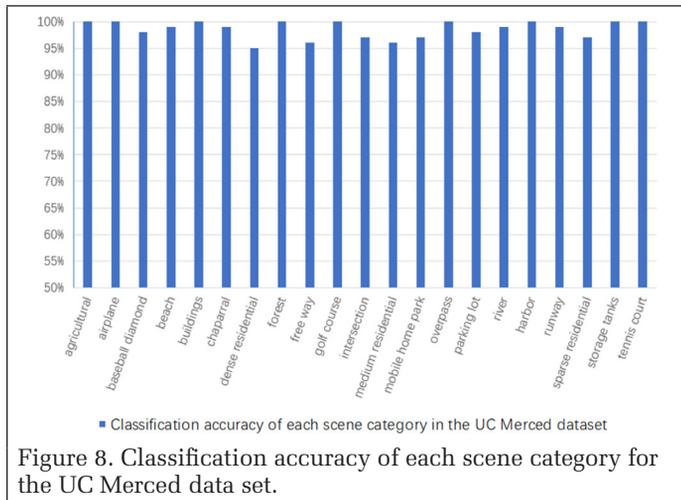


Figure 8. Classification accuracy of each scene category for the UC Merced data set.

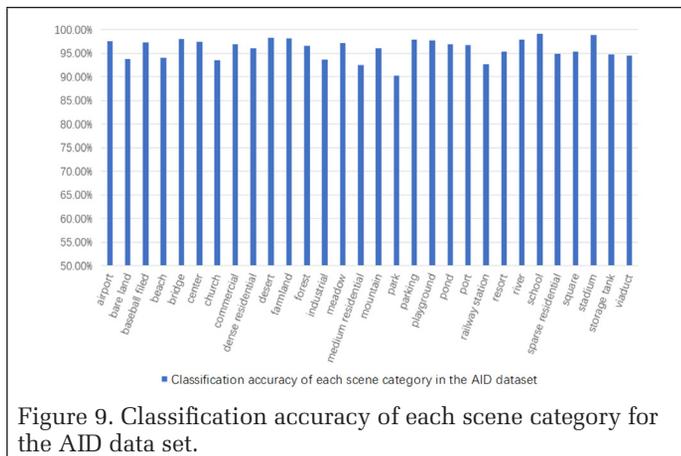


Figure 9. Classification accuracy of each scene category for the AID data set.

Table 6. Accuracy assessment (overall accuracy, %) of each classification method using the NWPU-RESISC45 data set.

Method	Train Ratio	Train Ratio
	80%	90%
Pretrained VGGNet-16	72.06	72.33
Pretrained Inception V3	79.10	79.70
Pretrained ResNet50	67.33	69.93
Pretrained VGGNet-19	76.58	77.33
Fine-tuned VGGNet-16	87.15	87.29
Fine-tuned Inception V3	84.90	86.38
Fine-tuned ResNet50	85.56	88.09
Fine-tuned VGGNet-19	88.29	89.11
HOG + SVM	35.88	40.22
LBP + SVM	29.20	32.16
SIFT + SVM	44.97	59.48
VGGNet-16_fc + SVM	81.85	82.10
Inception V3_fc + SVM	80.03	83.56
ResNet50_fc + SVM	79.97	85.45
VGGNet-19_fc + SVM	81.18	81.48
VGG16_conv + SVM	85.97	86.10
Inception V3_conv + SVM	85.47	85.89
VGG16_conv + BoW + SVM	82.15	83.02
Inception V3_conv + BoW + SVM	82.65	82.22
VGG16_conv + VLAD + SVM	87.33	87.79
ResNet + SIFT + SVM	89.58	90.25
ResNet + Inception V3 + SVM	88.93	89.22
Pretrained DenseNet201	81.22	82.57
Fine-tuned DenseNet201	90.36	92.18
Proposed method	94.16	95.09

BoW = XXXX; HOG = XXXX; LBP = XXXX; SIFT = scale-invariant feature transform; SVM = XXXX; VLAD = XXXX.

NWPU-RESISC45 data set is shown in Figure 10. The proposed method yielded a classification accuracy greater than 90% for most scene categories in the data set, and for aircraft, clouds, and forests it reached almost 100%. However, for some fine-grained scenes, such as medium and sparse residential areas, the classification accuracy was relatively poor, around 85%.

As shown in Tables 5 and Table 6, the proposed method performed well on the UC Merced, AID, and NWPU-RESISC45 data sets, with the best classification results on the UC Merced and AID data set. Compared to the method based on fine-tuned, pretrained ResNet50, the proposed method improved classification accuracy by 0.26% on the UC Merced data set. For the AID data set, the proposed method exhibited a classification accuracy of 95.83%, followed by the method based on fine-tuned, pretrained VGGNet. The proposed method also yielded the highest classification accuracy—94.16%—on the NWPU-RESISC45 data set. These results not only prove the effectiveness of the proposed method on small data sets but suggest that the method is also well suited to large and complex data sets. In addition, the effects of three different factors—the parameter  $K$  of the GMM, the feature-normalization methods, and the classifiers—on the classification accuracy for the NWPU-RESISC45 data set were explored. The training-set ratio in the subsequent experiments was 80%.

The value of  $K$  was set from 1 to 4 in this study. Table 7 shows that when the number of Gaussian components in the GMM is set to two, the proposed method can produce the best classification accuracy. To explore the effects of different normalization methods on the classification accuracy, three

Table 7. Classification accuracy with different  $K$  values for the NWPU-RESISC45 data set.

$K$	OA (%)	Number of Features
1	92.11	3840
2	94.16	7680
3	93.72	11 520
4	93.25	15 360

OA = overall accuracy.

Table 8. Classification results with different normalization methods.

	Normalization		
	No normalization	$\ell_2$ vector normalization	$\ell_2$ matrix normalization
OA (%)	91.99	92.56	94.16

OA = overall accuracy.

Table 9. Classification accuracy of different classifiers.

Classifier	OA (%)
RF	74.88
LR	87.21
SVM	91.15
Decision tree	73.38
MLP	85.43
Bayesian Gaussian	81.23
GBDT	80.65
Our classifier (decision tree)	85.80
Our classifier (Bayes)	89.90
Our classifier (LR)	94.16

GBDT = XXXX; LR = XXXX; MLP = XXXX; OA = overall accuracy; SVM = XXXX.

comparisons were carried out: no normalization,  $\ell_2$  vector normalization, and  $\ell_2$  matrix normalization. Table 8 suggests that feature normalization can greatly improve classification accuracy. Compared with  $\ell_2$  vector normalization,  $\ell_2$  matrix normalization produced better classification accuracy. This better classification performance should be attributed to the inclusion of the global features invariant to illumination or scale.

Table 9 shows the impact of different classifiers on the classification accuracy. The results suggest that an ensemble classifier stacking base classifiers exhibited better classification accuracy than individual classifiers. The effect of the number of iterations of the integrated classifier on the classification accuracy is shown in Figure 11. As the number of iterations increased, the classification accuracy improved. However, when the number of iterations reached around 45, the classification accuracy increased only slowly, eventually stabilizing. This slow increase or stagnation in classification accuracy may be due to the overfitting of base classifiers. Since the final classification results were voting results of all base classifiers, overfitting of base classifiers should have little effect on the final classification results.

## Conclusions

In recent years, achievements have been made in promoting CNNs for semantic-level scene classification of remotely sensed images. Despite great performance, existing methods based on CNNs for scene classification still have limitations, mainly because they generally use outputs from fully connected layers at the end of a CNN as classification features, ignoring local and spatial information that is vital in scene

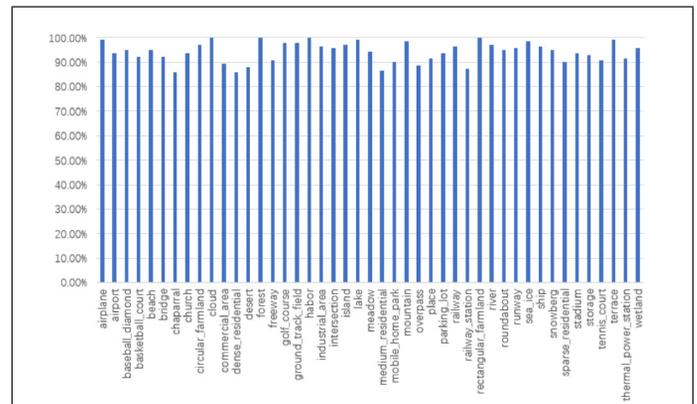


Figure 10. Classification accuracy of each scene class for the NWPU-RESISC45 data set.

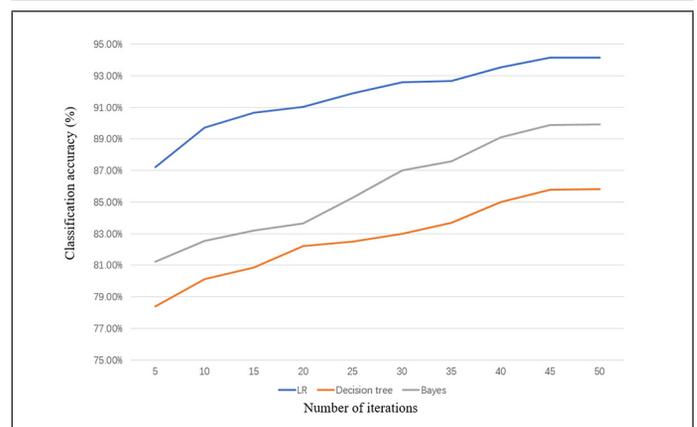


Figure 11. The effects on classification accuracy of the number of iterations in the AdaBoost.M1 algorithm.

classification of remotely sensed images. To address this issue, a scene-classification method of remotely sensed images based on a pretrained DenseNet combined with an ensemble classifier was proposed in this article. Although our work was motivated by the merits of DenseNet, obviously it illustrates more than a simple and direct application of DenseNet. Specifically, our proposed method not only took full advantage of DenseNet, with its distinguished capacity to learn hierarchical representation to describe the image scene, but also further strengthened the image-representation capability by using the pyramid structure and improved FV coding. At the same time, an ensemble classifier was built to overcome the limitation arising from a single classifier, which improved the classification performance. We synthetically considered local and spatial characteristics of remotely sensed data and their scale and rotation invariance, as well as the efficiency of the network. Results on three benchmark data sets showed that the proposed method exhibited the highest classification accuracy, particularly for the NSFU-RESISC45 data set, with a classification accuracy of 94.16%, which is superior to the state-of-the-art methods. Our future work will focus on exploiting other feature-coding strategies for optimization and exploring ensemble classifiers in depth, since the current one is relatively simple with low cost.

## Acknowledgments

The authors would like to thank the editors and anonymous reviewers for their valuable comments, which helped improve this work. The work of this article was supported by the National Key R&D Program of China (2018YFB0505401), the National Natural Science Foundation of China (41771452), and the Director Fund of the Institute of Remote Sensing and Digital Earth (Y5SJ1500CX).

## References

- Aptoula, E. 2014. Remote sensing image retrieval with global morphological texture descriptors. *IEEE Transactions on Geoscience and Remote Sensing* 52 (5):3023–3034.
- Bhagavathy, S. and B. S. Manjunath. 2006. Modeling and detection of geospatial objects using texture motifs. *IEEE Transactions on Geoscience and Remote Sensing* 44 (12):3706–3715.
- Blei, D. M., A. Y. Ng and M. I. Jordan. 2003. Latent Dirichlet allocation. *Journal of Machine Learning Research* 3:993–1022.
- Bosch, A., A. Zisserman and X. Muñoz. 2006. Scene classification via pLSA. Pages 517–530 in *Computer Vision—ECCV 2006*, held in Graz, Austria, 7–13 May 2006. Edited by A. Leonardis, H. Bischof and A. Pinz. Lecture Notes in Computer Science vol. 3954. Berlin: Springer.
- Bui, T. D., J. Shin, and T. Moon. 2019. Skip-connected 3D DenseNet for volumetric infant brain MRI segmentation. *Biomedical Signal Processing and Control* 54:101613.
- Castelluccio, M., G. Poggi, C. Sansone and L. Verdoliva. 2015. Land Use Classification in Remote Sensing Images by Convolutional Neural Networks. <<http://arxiv.org/abs/1508.00092>> Accessed DD Month YYYY.
- Chaib, S., H. Liu, Y. Gu and H. Yao. 2017. Deep feature fusion for VHR remote sensing scene classification. *IEEE Transactions on Geoscience and Remote Sensing* 55 (8):4775–4784.
- Chakraborty, D., S. Singh and D. Dutta. 2017. Segmentation and classification of high spatial resolution images based on Hölder exponents and variance. *Geo-spatial Information Science* 20 (1):39–45.
- Cheng, G., J. Han, L. Guo and T. Liu. 2015. Learning coarse-to-fine sparselets for efficient object detection and scene classification. Pages 1173–1181 in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, held in Boston, Mass., 7–12 June 2015. Edited by J. Editor. City, St.: IEEE.
- Cheng, G., J. Han, L. Guo, Z. Liu, S. Bu and J. Ren. 2015. Effective and efficient midlevel visual elements-oriented land-use classification using VHR remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing* 53 (8):4238–4249.
- Cheng, G., J. Han, and X. Lu. 2017. Remote sensing image scene classification: Benchmark and state of the art. *Proceedings of the IEEE* 105 (10):1865–1883.
- Cheng, G., J. Han, P. Zhou and L. Guo. 2014. Multi-class geospatial object detection and geographic image classification based on collection of part detectors. *ISPRS Journal of Photogrammetry and Remote Sensing* 98:119–132.
- Cheng, G., Z. Li, X. Yao, L. Guo and Z. Wei. 2017. Remote sensing image scene classification using bag of convolutional features. *IEEE Geoscience and Remote Sensing Letters* 14 (10):1735–1739.
- Cheng, G., C. Yang, X. Yao, L. Guo and J. Han. 2018. When deep learning meets metric learning: Remote sensing image scene classification via learning discriminative CNNs. *IEEE Transactions on Geoscience and Remote Sensing* 56 (5):2811–2821.
- Cheng, G., P. Zhou, and J. Han. 2016. Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing* 54 (12):7405–7415.
- Cheriyadat, A. M. 2013. Unsupervised feature learning for aerial scene classification. *IEEE Transactions on Geoscience and Remote Sensing* 52 (1):439–451.
- Chu, S. W., Y. Song, J. J. Zouo and W. Cai. 2019. Human pose estimation using deep convolutional Densenet hourglass network with intermediate points voting. Pages 594–598 in *2019 IEEE International Conference on Image Processing (ICIP)*, held in Taipei, Taiwan, 22–25 September 2019. Edited by J. Editor. City, St.: IEEE.
- Csurka, G., C. R. Dance, L. Fan, J. Willamowski and C. Bray. 2004. Visual categorization with bags of keypoints. Pages 1–2 in *Workshop on Statistical Learning in Computer Vision, ECCV*, held in Prague, Czech Republic, DD–DD Month YYYY. Edited by J. Editor. City, St.: Publisher.
- Dede, M. A., E. Aptoula and Y. Genc. 2019. Deep network ensembles for aerial scene classification. *IEEE Geoscience and Remote Sensing Letters* 16 (5):732–735.
- Dezaki, F. T., Z. Liao, C. Luong, H. Girgis, N. Dhungel, A. H. Abdi, D. Behnami, K. Gin, R. Rohling, P. Abolmaesumi and T. Tsang. 2019. Cardiac phase detection in echocardiograms with densely gated recurrent neural networks and global extrema loss. *IEEE Transactions on Medical Imaging* 38 (8):1821–1832.
- dos Santos, J. A., O.A.B. Penatti and R. da Silva Torres. 2010. Evaluating the potential of texture and color descriptors for remote sensing image retrieval and classification. Pages 203–208 in *VISAPP (2)*, held in City, St., DD–DD Month YYYY. Edited by J. Editor. City, St.: Publisher.
- Faraji, M. and J. Shanbehzadeh. 2015. Bag-of-Visual-Words, its detectors and descriptors; a survey in detail. *Advances in Computer Science: An International Journal* 4 (2):8–20.
- Flores, E., M. Zortea and J. Scharcanski. 2019. Dictionaries of deep features for land-use scene classification of very high spatial resolution images. *Pattern Recognition* 89:32–44.
- Gao, B.-B., X.-S. Wei, J. Wu and W. Lin. 2015. Deep Spatial Pyramid: The Devil Is Once Again in the Details. <<https://arxiv.org/abs/1504.05277>> Accessed DD Month YYYY.
- Girshick, R., J. Donahue, T. Darrell and J. Malik. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. Pages 580–587 in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, held in Columbus, Ohio, 23–28 June 2014. Edited by J. Editor. City, St.: IEEE.
- Gu, J., Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai and T. Chen. 2018. Recent advances in convolutional neural networks. *Pattern Recognition* 77:354–377.
- Hao, W. and Z. Zhang. 2019. Spatiotemporal distilled dense-connectivity network for video action recognition. *Pattern Recognition* 92:13–24.

- Haralick, R. M., K. Shanmugam and I. Dinstein. 1973. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics SMC-3* (6):610–621.
- He, K., X. Zhang, S. Ren and J. Sun. 2015. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37 (9):1904–1916.
- He, K., X. Zhang, S. Ren and J. Sun. 2016. Deep residual learning for image recognition. Pages 770–778 in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, held in Las Vegas, Nev., 27–30 June 2016. Edited by J. Editor. City, St.: IEEE.
- He, X., Y. Yang, B. Shi and X. Bai. 2019. VD-SAN: Visual-Densely Semantic Attention Network for image caption generation. *Neurocomputing* 328:48–55.
- Heipke, C. and F. Rottensteiner. 2020. Deep learning for geometric and semantic tasks in photogrammetry and remote sensing. *Geospatial Information Science* 23 (1):10–19.
- Hu, F., G.-S. Xia, J. Hu and L. Zhang. 2015. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. *Remote Sensing* 7 (11):14680–14707.
- Hu, F., W. Yang, J. Chen and H. Sun. 2013. Tile-level annotation of satellite images using multi-level max-margin discriminative random field. *Remote Sensing* 5 (5):2275–2291.
- Huang, G., Z. Liu, L. Van Der Maaten and K. Q. Weinberger. 2017. Densely connected convolutional networks. Pages 4700–4708 in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, held in Honolulu, Hawaii, 21–26 July 2017. Edited by J. Editor. City, St.: Publisher.
- Huang, L., C. Chen, W. Li and Q. Du. 2016. Remote sensing image scene classification using multi-scale completed local binary patterns and Fisher vectors. *Remote Sensing* 8 (6):483.
- Huo, L.-J., B. He and D.-B. Zhou. 2017. A destriping method with multi-scale variational model for remote sensing images. *Optics and Precision Engineering* 25 (1):198–207.
- Jégou, H., M. Douze, C. Schmid and P. Pérez. 2010. Aggregating local descriptors into a compact image representation. Pages 3304–3311 in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, held in San Francisco, Calif., 13–18 June 2010. Edited by J. Editor. City, St.: IEEE.
- Jia, Y., E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama and T. Darrell. 2014. Caffe: Convolutional architecture for fast feature embedding. Pages 675–678 in *MM '14: Proceedings of the 22nd ACM International Conference on Multimedia*, held in Orlando, Fla., November 2014. Edited by J. Editor. New York: Association for Computing Machinery.
- Khened, M., V. A. Kollerathu and G. Krishnamurthi. 2019. Fully convolutional multi-scale residual DenseNets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers. *Medical Image Analysis* 51:21–45.
- Krizhevsky, A., I. Sutskever and G. E. Hinton. 2012. ImageNet classification with deep convolutional neural networks. Pages 1097–1105 in *Advances in Neural Information Processing Systems 25*, held in Lake Tahoe, Nev., 3–6 December 2012. Edited by P. Bartlett. City, St.: Neural Information Processing Systems Foundation.
- Kusumaningrum, R., H. Wei, R. Manurung and A. Murni. 2014. Integrated visual vocabulary in latent Dirichlet allocation-based scene classification for IKONOS image. *Journal of Applied Remote Sensing* 8 (1):083690.
- Lazebnik, S., C. Schmid and J. Ponce. 2006. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. Pages 2169–2178 in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, held in New York, N.Y., 17–22 June 2006. Edited by J. Editor. City, St.: IEEE.
- Li, G., C. Zhang, M. Wang, F. Gao and X. Zhang. 2018. Scene classification of high-resolution remote sensing image using transfer learning with multi-model feature extraction framework. Pages 238–251 in *Image and Graphics Technologies and Applications*, held in Beijing, China, 8–10 April 2018. Edited by Y. Wang, Z. Jiang and Y. Peng. Communications in Computer and Information Science vol. 875. Singapore: Springer.
- Li, H., H. Gu, Y. Han and J. Yang. 2010. Object-oriented classification of high-resolution remote sensing imagery based on an improved colour structure code and a support vector machine. *International Journal of Remote Sensing* 31 (6):1453–1470.
- Li, X., X. Yao and Y. Fang. 2018. Building-A-Nets: Robust building extraction from high-resolution remote sensing images with adversarial networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 11 (10):3680–3687.
- Lienou, M., H. Maitre and M. Datcu. 2010. Semantic annotation of satellite images using latent Dirichlet allocation. *IEEE Geoscience and Remote Sensing Letters* 7 (1):28–32.
- Liu, Q., R. Hang, H. Song and Z. Li. 2018. Learning multiscale deep features for high-resolution satellite image scene classification. *IEEE Transactions on Geoscience and Remote Sensing* 56 (1):117–126.
- Liu, Y., Y. Zhong and Q. Qin. 2018. Scene classification based on multiscale convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing* 56 (12):7109–7121.
- Lowe, D. G. 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60:91–110.
- Lu, X., X. Zheng and Y. Yuan. 2017. Remote sensing scene classification by unsupervised representation learning. *IEEE Transactions on Geoscience and Remote Sensing* 55 (9):5148–5157.
- Luus, F.P.S., B. P. Salmon, F. van den Bergh and B.T.J. Maharaj. 2015. Multiview deep learning for land-use classification. *IEEE Geoscience and Remote Sensing Letters* 12 (12):2448–2452.
- Mahendran, A. and A. Vedaldi. 2015. Understanding deep image representations by inverting them. Pages 5188–5196 in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, held in Boston, Mass., 7–12 June 2015. Edited by J. Editor. City, St.: IEEE.
- Marmanis, D., M. Datcu, T. Esch and U. Stilla. 2016. Deep learning earth observation classification using ImageNet pretrained networks. *IEEE Geoscience and Remote Sensing Letters* 13 (1):105–109.
- Minetto, R., M. P. Segundo and S. Sarkar. 2019. Hydra: An ensemble of convolutional neural networks for geospatial land classification. *IEEE Transactions on Geoscience and Remote Sensing* 57 (9):6530–6541.
- Negrel, R., D. Picard and P.-H. Gosselin. 2014. Evaluation of second-order visual features for land-use classification. Pages 1–5 in *2014 12th International Workshop on Content-Based Multimedia Indexing (CBMI)*, held in Klagenfurt, Austria, 18–20 June 2014. Edited by J. Editor. City, St.: IEEE.
- Newsam, S., L. Wang, S. Bhagavathy and B. S. Manjunath. 2004. Using texture to analyze and manage large collections of remote sensed image and video data. *Applied Optics* 43 (2):210–217.
- Nogueira, K., O. A. B. Penatti and J.A. dos Santos. 2017. Towards better exploiting convolutional neural networks for remote sensing scene classification. *Pattern Recognition* 61:539–556.
- Penatti, O. A. B., K. Nogueira and J.A. dos Santos. 2015. Do deep features generalize from everyday objects to remote sensing and aerial scenes domains? Pages 44–51 in *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, held in Boston, Mass., 7–12 June 2015. Edited by J. Editor. City, St.: IEEE.
- Perronnin, F., J. Sánchez and T. Mensink. 2010. Improving the Fisher kernel for large-scale image classification. Pages 143–156 in *Computer Vision—ECCV 2010*, held in Heraklion, Greece, 5–11 September 2010. Edited by K. Daniilidis, P. Maragos and N. Paragios. Lecture Notes in Computer Science vol. 6314. Berlin: Springer.

- Risojevi, V. and Z. Babi. 2011. Aerial image classification using structural texture similarity. Pages 190–195 in *2011 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, held in Bilbao, Spain, 14–17 December 2011. Edited by J. Editor. City, St.: IEEE.
- Russakovsky, O., J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg and F.-F. Li. 2015. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision* 115:211–252.
- Sánchez, J., F. Perronnin, T. Mensink and J. Verbeek. 2013. Image classification with the Fisher vector: Theory and practice. *International Journal of Computer Vision* 105:222–245.
- Scott, G. J., M. R. England, W. A. Starms, R. A. Marcum and C. H. Davis. 2017. Training deep convolutional neural networks for land-cover classification of high-resolution imagery. *IEEE Geoscience and Remote Sensing Letters* 14 (4):549–553.
- Sermanet, P., D. Eigen, X. Zhang, M. Mathieu, R. Fergus and Y. LeCun. 2013. OverFeat: Integrated Recognition, Localization and Detection Using Convolutional Networks. <<https://arxiv.org/abs/1312.6229>> Accessed DD Month YYYY.
- Sheng, G., W. Yang, T. Xu and H. Sun. 2012. High-resolution satellite scene classification using a sparse coding based multiple feature combination. *International Journal of Remote Sensing* 33 (8):2395–2412.
- Simonyan, K. and A. Zisserman. 2015. Very deep convolutional networks for large-scale image recognition. <<https://arxiv.org/abs/1409.1556>> Accessed DD Month YYYY.
- Stricker, M. A. and M. Orengo. 1995. Similarity of color images. *Proceedings of SPIE* 2420:381–392.
- Szegedy, C., W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich. 2015. Going deeper with convolutions. Pages 1–9 in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, held in Boston, Mass., 7–12 June 2015. Edited by J. Editor. City, St.: IEEE.
- Taigman, Y., M. Yang, M. Ranzato and L. Wolf. 2014. DeepFace: Closing the gap to human-level performance in face verification. Pages 1701–1708 in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, held in Columbus, Ohio, 23–28 June 2014. Edited by J. Editor. City, St.: IEEE.
- Tao, Y., M. Xu, Z. Lu and Y. Zhong. 2018. DenseNet-based depth-width double reinforced deep learning neural network for high-resolution remote sensing image per-pixel classification. *Remote Sensing* 10 (5):779.
- Wan, L., N. Liu, Y. Guo, H. Huo and T. Fang. 2017. Local feature representation based on linear filtering with feature pooling and divisive normalization for remote sensing image classification. *Journal of Applied Remote Sensing* 11 (1):016017.
- Wang, H., Y. Shen, S. Wang, T. Xiao, L. Deng, X. Wang and X. Zhao. 2019. Ensemble of 3D densely connected convolutional network for diagnosis of mild cognitive impairment and Alzheimer's disease. *Neurocomputing* 333:145–156.
- Wang, J., C. Luo, H. Huang, H. Zhao and S. Wang. 2017. Transferring pre-trained deep CNNs for remote scene classification with general features learned from linear PCA network. *Remote Sensing* 9 (3):225.
- Wang, L., L. Qiu, W. Sui and C. Pan. 2018. Reconstructed Densenets for image super-resolution. Pages 3558–3562 in *2018 25th IEEE International Conference on Image Processing (ICIP)*, held in Athens, Greece, 7–10 October 2018. Edited by J. Editor. City, St.: IEEE.
- Xia, G.-S., J. Hu, F. Hu, B. Shi, X. Bai, Y. Zhong, L. Zhang and X. Lu. 2017. AID: A benchmark data set for performance evaluation of aerial scene classification. *IEEE Transactions on Geoscience and Remote Sensing* 55 (7):3965–3981.
- Yang, Y. and S. Newsam. 2008. Comparing SIFT descriptors and gabor texture features for classification of remote sensed imagery. Pages 1852–1855 in *2008 15th IEEE International Conference on Image Processing*, held in San Diego, Calif., 12–15 October 2008. Edited by J. Editor. City, St.: IEEE.
- Yang, Y. and S. Newsam. 2010. Bag-of-visual-words and spatial extensions for land-use classification. Pages 270–279 in *GIS '10: Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, held in San Jose, Calif., November 2010. Edited by J. Editor. New York: Association for Computing Machinery.
- Yang, Y. and S. Newsam. 2011. Spatial pyramid co-occurrence for image classification. Pages 1465–1472 in *2011 International Conference on Computer Vision*, held in Barcelona, Spain, 6–13 November 2011. City, St.: IEEE.
- Yosinski, J., J. Clune, A. Nguyen, T. Fuchs and H. Lipson. 2015. Understanding Neural Networks Through Deep Visualization. <<https://arxiv.org/abs/1506.06579>> Accessed DD Month YYYY.
- Yu, H., J. Wang, Y. Bai, W. Yang and G.-S. Xia. 2018. Analysis of large-scale UAV images using a multi-scale hierarchical representation. *Geo-spatial Information Science* 21 (1):33–44.
- Yu, W., K. Yang, Y. Bai, H. Yao and Y. Rui. 2014. DNN flow: DNN feature pyramid based image matching. Pages PP–PP in *BMVC*, held in City, St., DD–DD Month YYYY. Edited by J. Editor. City, St.: Publisher.
- Zhang, C., G. Li and S. Du. 2019. Multi-scale dense networks for hyperspectral remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing* 57 (11):9201–9222.
- Zhang, F., B. Du and L. Zhang. 2016. Scene classification via a gradient boosting random convolutional network framework. *IEEE Transactions on Geoscience and Remote Sensing* 54 (3):1793–1802.
- Zhang, K., Y. Guo, X. Wang, J. Yuan, Z. Ma and Z. Zhao. 2019. Channel-wise and feature-points reweights Densenet for image classification. Pages 410–414 in *2019 IEEE International Conference on Image Processing (ICIP)*, held in Taipei, Taiwan, 22–25 September 2019. Edited by J. Editor. City, St.: IEEE.
- Zhang, R., L. Zhao, W. Lou, J. M. Abrigo, V.C.T. Mok, W.C.W. Chu, D. Wang and L. Shi. 2018. Automatic segmentation of acute ischemic stroke from DWI using 3-D fully convolutional DenseNets. *IEEE Transactions on Medical Imaging* 37 (9):2149–2160.
- Zhang, W., P. Tang and L. Zhao. 2019. Remote sensing image scene classification using CNN-CapsNet. *Remote Sensing* 11 (5):494.
- Zhang, Z., X. Liang, X. Dong, Y. Xie and G. Cao. 2018. A sparse-view CT reconstruction method based on combination of DenseNet and deconvolution. *IEEE Transactions on Medical Imaging* 37 (6):1407–1417.
- Zhao, B., B. Huang and Y. Zhong. 2017. Transfer learning with fully pretrained deep convolution networks for land-use classification. *IEEE Geoscience and Remote Sensing Letters* 14 (9):1436–1440.
- Zhao, L., P. Tang and L. Huo. 2014a. A 2-D wavelet decomposition-based bag-of-visual-words model for land-use scene classification. *International Journal of Remote Sensing* 35 (6):2296–2310.
- Zhao, L.-J., P. Tang and L.-Z. Huo. 2014b. Land-use scene classification using a concentric circle-structured multiscale bag-of-visual-words model. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7 (12):4620–4631.
- Zheng, X., Y. Yuan and X. Lu. 2019. A deep scene representation for aerial scene classification. *IEEE Transactions on Geoscience and Remote Sensing* 57 (7):4799–4809.
- Zhong, Y., M. Cui, Q. Zhu and L. Zhang. 2015. Scene classification based on multifeature probabilistic latent semantic analysis for high spatial resolution remote sensing images. *Journal of Applied Remote Sensing* 9 (1):095064.
- Zhou, F., X. Li and Z. Li. 2018. High-frequency details enhancing DenseNet for super-resolution. *Neurocomputing* 290:34–42.

# Digital Elevation Model Technologies and Applications: The DEM Users Manual, 3<sup>rd</sup> Edition

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

To order, visit  
<https://www.asprs.org/dem>

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  
*David F. Maune, Hans Karl Heidemann, Stephen M. Kopp, and Clayton A. Crawford*
  2. Vertical Datums  
*Dru Smith*
  3. Standards, Guidelines & Specifications  
*David F. Maune*
  4. The National Elevation Dataset (NED)  
*Dean B. Gesch, Gayla A. Evans, Michael J. Oimoen, and Samantha T. Arundel*
  5. The 3D Elevation Program (3DEP)  
*Jason M. Stoker, Vicki Lukas, Allyson L. Jason, Diane F. Eldridge, and Larry J. Sugarbaker*
  6. Photogrammetry  
*J. Chris McGlone and Scott Arko*
  7. IfSAR  
*Scott Hensley and Lorraine Tighe*
  8. Airborne Topographic Lidar  
*Amar Nayegandhi and Joshua Nimetz*
  9. Lidar Data Processing  
*Joshua M. Novac*
  10. Airborne Lidar Bathymetry  
*Jennifer Wozencraft and Amar Nayegandhi*
  11. Sonar  
*Guy T. Noll and Douglas Lockhart*
  12. Enabling Technologies  
*Bruno M. Scherzinger, Joseph J. Hutton, and Mohamed M.R. Mostafa*
  13. DEM User Applications  
*David F. Maune*
  14. DEM User Requirements & Benefits  
*David F. Maune*
  15. Quality Assessment of Elevation Data  
*Jennifer Novac*
- Appendix A. Acronyms  
Appendix B. Definitions  
Appendix 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 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

---

## PRICING

Student (must submit copy of Student ID)	\$50 +S&H
ASPRS Member	\$80 +S&H
Non-member	\$100 +S&H
E-Book (only available in the Amazon Kindle store)	\$85

---

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](http://asprs.org)

ASPRS