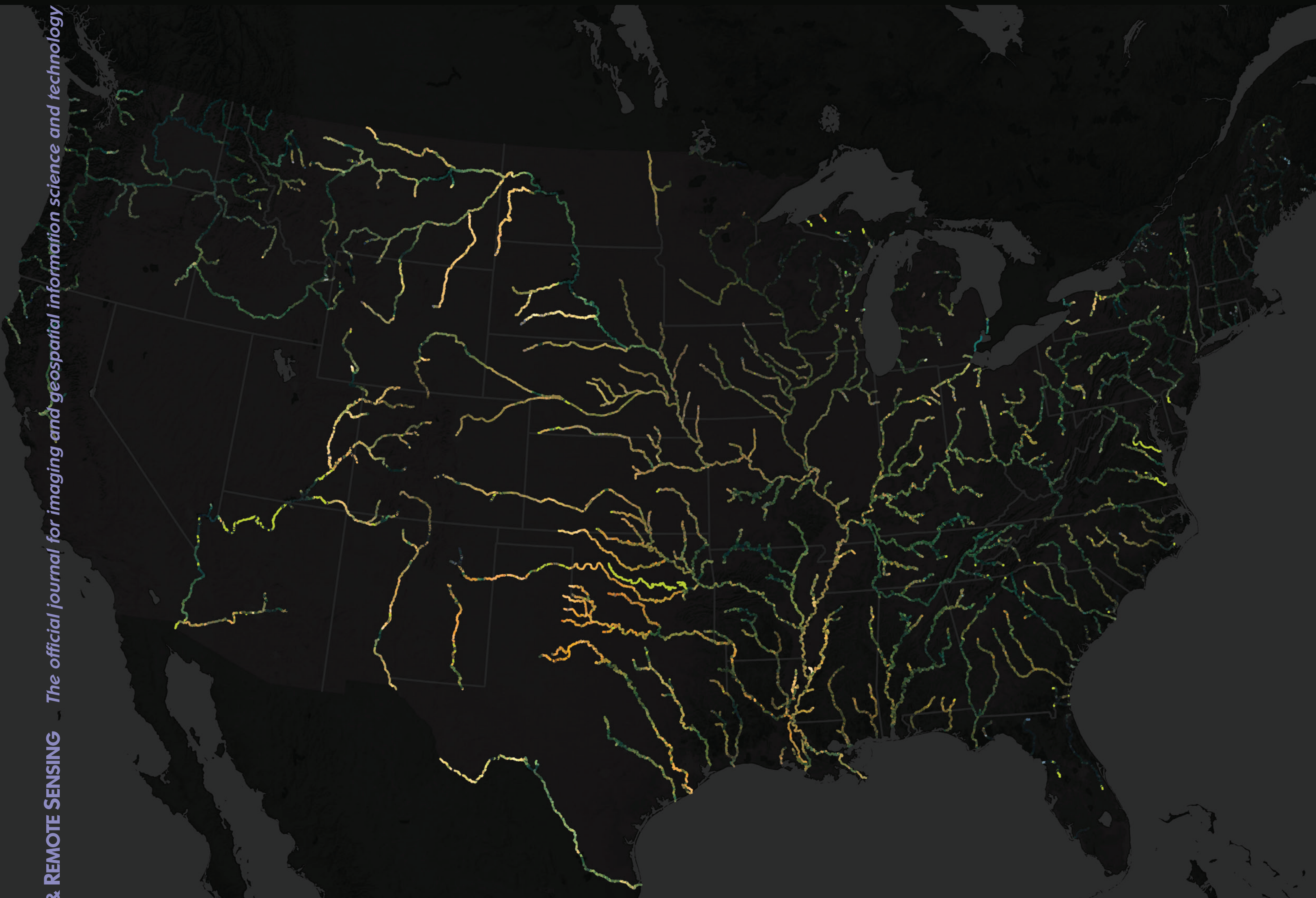
PE&RS

January 2022 Volume 88, Number 1



PHOTOGRAMMETRIC ENGINEERING & REMOTE SENSING

The official journal for imaging and geospatial information science and technology

ASPRS 2022 Annual Conference at Geo Week in Denver, CO, February 6 - 8, 2022

In 2022, ASPRS will transition back to a traditional face-to-face conference format as part of Geo Week 2022. However, we realized that our virtual ASPRS conference in 2021 reached many attendees and presenters who, regardless of COVID, would not be able to take part in Geo Week due to travel or cost constraints. Therefore, we are holding our 2022 ASPRS Annual Conference in two parts.

The February 6-8, 2022, conference in Denver, Colorado will be a live, in-person event. It will not be broadcast or recorded. Attendees can participate in:

- ASPRS technical sessions consisting of individual presentations, panel discussions, and a poster gallery
- ASPRS 2-hour and 4-hour workshops
- Shared Geo Week exhibit hall and social networking functions
- ASPRS Board of Directors, Council, Division, and Committee meetings
- ASPRS Annual Business Meeting, including
  - o Presentation of ASPRS Awards and Scholarships
  - o Installation of Officers
- Past Presidents and ASPRS Foundation Board of Trustees meetings (by invitation)
- Student and Early Career Professional Mentoring and Networking

ASPRS 2022 Virtual Technical Program online, March 21 - 25, 2022

A live webinar event consisting of oral presentations, panel discussions, and an online poster gallery. This event will be recorded and made available on-demand.

For information on registering and presenting visit **https://my.asprs.org/2022conference**

For additional information or questions, contact programs@asprs.org.

Photo by Andrew Coop on Unsplash

ASPRS 2022 Annual Conference Workshops

At Geo Week | February 6-8, 2022 | Denver, CO, USA

Aerial Triangulation and Data Processing for the Unmanned Aerial System (UAS)

Dr. Qassim Abdullah, PhD, CP, PLS, Vice President/Chief Scientist, *Woolpert, Inc.*
Dr. Riadh Munjy, Professor, *California State University*, Fresno

This workshop teaches participants to successfully design, plan and execute an aerial mission using unmanned aerial systems (UAS) and GPS-based aerial triangulation, including flight planning, ground control placement, camera calibration, and product generation. Participants will be introduced to mathematical basis of simultaneous bundle block adjustment.

February 6th, 8:00 AM to 12:15 PM Cost: $250

Fundamentals of Image Analysis in Google Earth Engine

Dr. Ge (Jeff) Pu, PhD, CMS, *NOAA Great Lakes Environmental Research Laboratory and Cleveland Water Alliance*

This workshop will provide an interactive exploration of Google Earth Engine (GEE) capabilities with examples of projects demonstrating the use of GEE in education undergraduate research and outreach followed by more advanced topics of GEE that includes image processing widget use and app building using an API based coding environment. Each participant will get at least 4 GEE activities for classroom use and several GEE API scripts that the participants can modify for their own use. The 4 activities will include image classification and accuracy assessment, image shadow detection and removal, time series analysis, advanced script sharing and app development.

Special Requirements: Laptop with WiFi Capability.

February 6th, 8:00 AM to 12:15 PM Cost: $250

Preparation for ASPRS Certification

Youssef Kaddoura, PhD Candidate/Research Scientist, *University of Florida*

In this workshop, attendees will review fundamental knowledge areas covered by ASPRS certification exams (photogrammetry, remote sensing, GIS, lidar, and UAS).

February 6th, 8:00 AM to 12:15 PM Cost: $250

Airborne Topobathy Lidar — Principles and Applications

Amar Nayeghandi, Senior Vice President, *Dewberry*
Nick Kules, Senior Geospatial Technology Manager, *Dewberry*
Christopher Parrish, Associate Professor, *Oregon State University*

Airborne laser (or lidar) bathymetry (ALB) is a technique for measuring depths of nearshore coastal waters, lakes, and rivers from a low-altitude aircraft, typically using using a scanning, pulsed laser beam. Based on three decades of operations, ALB has proven to be an accurate, cost-effective, rapid, safe, and flexible method for surveying in shallow water and on coastlines where sonar systems are less efficient and can even be dangerous to operate. This seminar will cover the principles and applications of this technology, including an overview of the history of this technology and an overview of the sensors available today.

February 6th, 8:00 AM to 12:15 PM Cost: $250

Practical Approach to Using the ASPRS Positional Accuracy Standards for Digital Geospatial Data

Dr. Qassim Abdullah, PhD, CP, PLS, Vice President/Chief Scientist, *Woolpert, Inc.*
Claire Kiedrowski, Executive Director, *Maine Library of Geographic Information*

This workshop provides an in-depth look at the ASPRS Positional Accuracy Standards to categorize positional accuracy of products derived from digital aerial cameras, manned and unmanned aerial systems, and all types of lidar including terrestrial, mobile, and airborne. The workshop will explain the basis for each accuracy measure adopted in the standards. Instructors will demonstrate practical application of these standards.

February 6th, 12:45 PM to 5:00 PM Cost: $250

Machine and Deep Learning Image Classification using ArcGIS Pro

Dr. Amr Abd-Elrahman, Associate Professor, *University of Florida*

This workshop introduces pixel- and object-based image classification using traditional machine learning algorithms as well as deep learning semantic segmentation using the UNet Model, including hands on-activities in ArcGIS Pro.

Special Requirements: Laptop with WiFi Capability.

February 6th, 12:45 PM to 5:00 PM Cost: $250

Planning for the New National Spatial Reference System and Vertical Datum

Barry Miller, Applied Researcher, *USGS*
Josh Nimetz, Senior Elevation Product Lean, *USGS*

The USGS National Map is the Nation's source for topographic, hydrographic, and cartographic geospatial data. These National datasets may span collection periods of many decades and exist in a variety of different coordinate reference systems. This workshop will focus on planning discussions underway within the USGS National Geospatial Program and anticipated difficulties in transforming existing data holdings to the new reference system and vertical datum.

February 6th, 12:45 PM to 2:45 PM Cost: $175

Lidar Survey with UAS: Project Planning and Flight Operations

Ryan Kelly, Senior Geospatial Manager, *Halff Associates*

This ASPRS workshop will use real world use cases to compare lidar systems to determine best fit for various applications, demonstrate proven best flight practices for collecting high accuracy data, explore control placement and targeting, and communicate standards to surveyors and customers.

February 6th, 12:45 PM to 2:45 PM Cost: $175

USGS 3DEP Data Validation

Dr. Milena Janiec, Applied Researcher, *USGS*

In support of 3DEP and the elevation theme of The National Map, USGS performs data validation on topographic data collected with remote sensing technologies, primarily airborne lidar. This workshop will focus on the USGS policies and processes related to validation of airborne lidar for 3DEP.

February 6th, 3:00 PM to 5:00 PM Cost: $175

Lidar Survey with UAS: Data Processing and Product Validation

Rusty Steel, Geospatial Director, *Halff Associates*

This workshop will demystify lidar data processing, identify common mistakes, demonstrate quality inspection methods, and show how to apply ASPRS Positional Accuracy Standards to validate results.

February 6th, 3:00 PM to 5:00 PM Cost: $175

To have your press release published in *PE&RS*, contact Rae Kelley, rkelley@asprs.org.

ANNOUNCEMENTS

GeoCue is excited to announce that we are, once again, expanding our line of True View 3D imaging sensors (3DIS®). The True View 435 is our next-generation topography/wire grade lidar/imaging sensor. Featuring a 16 beam Hesai PandarXT LIDAR unit, world class Trimble Applanix APX-15 Position/Orientation System (POS) and dual GeoCue mapping cameras, the True View 435 is the highest performance 3D Imaging system available in its price class.

The True View 435 has a pulse repetition rate of 320,000 out-going pulses per second with an in-track field of view of +/- 15° of nadir. The sensor can detect up to two return pulses ("echoes") per outbound pulse. This provides excellent look angles for collecting power pylons/towers as well as superb, multiangle penetrating power in vegetated areas. Like all True View 3DIS, the True View 435 features dual oblique cameras, providing a cross-track image field of view of 120°. Similar to our industry standard True View 515, the True View 435 has sufficient sensitivity to detect distribution wires at an altitude of 75 m.

The True View 435 includes complete post-processing software featuring GeoCue's unique on-the-fly lidar point colorization algorithm, producing stunning 3D colorized point clouds. Through our teaming with BayesMap Solutions LLC, StripAlign for EVO (SAfE) is available as a purchased optional module or under a pay-as-you-use ("metered") plan. SAfE performs geometric correction for those occasional large projects that might exhibit dynamic geometric errors. Also available in the post-processing suite is Metashape for EVO (MfE), used in generating image orthomosaics based on lidar surfaces.

The True View 435 is upgradable to a True View 515, should you encounter situations where higher point densities are required.

The True View 435 is immediately available from GeoCue or GeoCue authorized resellers.

With the goal of giving surveyors a better understanding of the topographic data captured by drone mapping sensors, Virtual Surveyor has unveiled Profile View functionality in Version 8.4 of its popular surveying software. Profile View enables users to generate an elevation profile simply by drawing an onscreen traverse across any part of the data set created from drone imagery or lidar point clouds.

"For comprehensive understanding of the terrain, surveyors need to view their elevation data in 2D, 3D and in profile," said Tom Op 't Eyndt, Virtual Surveyor CEO. "The Virtual Surveyor software now offers all three types of viewing so users can look at their data from any angle and perspective."

Virtual Surveyor is a robust surveying software that bridges the gap between drone photogrammetric processing applications and engineering design packages, enabling surveyors to derive topographic information from drone data needed by engineers for construction, mining, and excavation projects. The software presents an interactive onscreen environment with drone orthophotos, digital surface models (DSMs), and/or lidar point clouds where users generate CAD models, create cut-and-fill maps, and calculate volume reports.

The Profile View allows users to draw straight or curved lines to cut across the terrain surface or follow an irregular feature, such as a road. The elevation profile is displayed in a new window that opens on screen. Profile View functionality will be valuable for surveyors working in any application related to construction, surface mining, landfill, and other types of excavation.

Current subscribers to Virtual Surveyor will see their software updated to Version 8.4 automatically. To start a free 14-day trial of Virtual Surveyor and to view details of the Valley, Ridge and Peak pricing plans, visit www.virtual-surveyor.com.

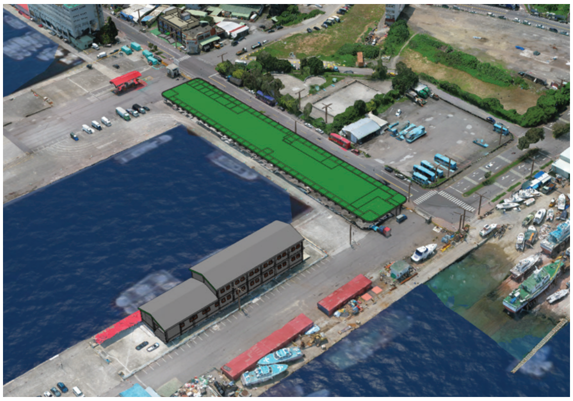
Applanix, a Trimble Company (NASDAQ: TRMB), and the National Oceanic and Atmospheric Administration (NOAA) recently collaborated to provide critical information to first responders in the wake of Hurricanes Henri and Ida. Applanix's high-accuracy direct georeferencing (DG) technology enabled NOAA to quickly collect aerial mapping imagery to provide valuable disaster remediation information to first responders, and demonstrate the value of cutting-edge mapping technology in preparing for and responding to emergency situations such as hurricanes, tornadoes and other disasters.

Within hours of Hurricanes Henri and Ida making landfall, NOAA's National Geodetic Survey collected post-storm imagery using the latest generation Digital Sensor System (DSS). The sixth generation DSS, designed and manufactured for Applanix by Lead'Air, is the most powerful to date, thanks to several new features introduced within the solution:

- Simultaneous full color and near-infrared image capture using high-performance Phase One iXM 100 MP NIR and 150 MP RGB cameras,
- Option to fly the cameras in wide coverage oblique or traditional overhead (straight line down) mode for mapping with uninterrupted measurement,
- Embedded Trimble AP60 global navigation satellite system-inertial (GNSS-inertial) OEM DG solution for mapping without the need for ground control or aerial triangulation,
- Applanix POSPac™ post-processing software featuring

continued on page 8

PE&RS
PHOTOGRAMMETRIC ENGINEERING & REMOTE SENSING
The official journal for imaging and geospatial information science and technology
January 2022 Volume 88 Number 1



COLUMNS

ANNOUNCEMENTS

DEPARTMENTS

See the Cover Description on Page 4

facebook.com/ASPRS.org linkedin.com/groups/2745128 twitter.com/ASPRSorg youtube.com/user/ASPRS

COVER DESCRIPTION

Much like the sky, rivers are rarely painted one color. Across the world, they appear in shades of yellow, green, blue, and brown. Subtle changes in the environment can alter the color of rivers, though, shifting them away from their typical hues. New research shows the dominant color has changed in about one-third of large rivers in the continental United States over the past 35 years.

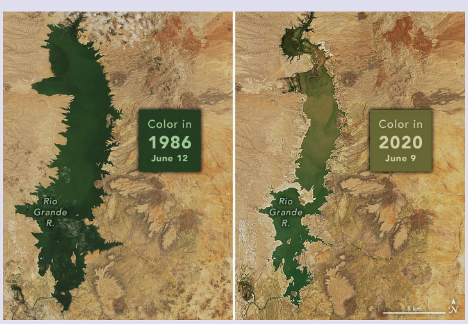
“Changes in river color serve as a first pass that tell us something is going on nearby,” said John Gardner, the study’s lead author and a hydrologist at the University of Pittsburgh. “There are a lot of details to parse out on what is causing those changes, though.”

The figure above shows data from the first map of river color for the contiguous United States. The rivers are colored as they would approximately appear to our eye. Gardner and colleagues built the map from 234,727 images collected by Landsat satellites between 1984 and 2018. The dataset includes 67,000 miles (100,000 kilometers) of waterways of at least 200 feet (60 meters) wide. Around 56 percent of rivers were dominantly yellow over the course of the study and 38 percent were dominantly green. The team has released an interactive map where the public can further investigate color trends in individual rivers.

It is not unusual for rivers to change colors, explained Gardner. They change all the time because of fluctuations in flow, concentrations of sediments, and the amount of dissolved organic matter or algae in the water. For example, yellow-tinted rivers are typically sediment-laden but low in algae. Blue water, which is usually easier to see through, has little algae and sediment. Green water usually has algae as its dominant feature.

“We are seeing an increase in the frequency of color changes,” said Gardner. In the study, the team found around 21 percent of rivers became greener, most commonly in the western United States. Around 12 percent of the rivers shifted towards yellow, many in the eastern United States.

The scientists found that the most extreme examples are often found near man-made reservoirs. In fact, the rivers with the fastest rate of color change were twice as likely to be located within 15 miles (25 kilometers) upstream or downstream of a dam and within the boundaries of an urban area.



The images to the left show color changes from 1986 (Landsat 5) to 2020 (Landsat 8) along the Rio Grande River near the Elephant Butte Reservoir in New Mexico. Gardner explained that changes in a reservoir’s surface area can affect river color. When reservoirs contract, the upstream end of reservoirs become sediment-laden rivers again. Gardner is currently working to estimate suspended sediment concentrations based on the Landsat dataset. The goal is to explore how human activities, such as construction of dams or land use, may be affecting sediment loads.

From his own observations, Gardner also noticed more occurrences of algal blooms in rivers. In 2015, an algal bloom stretched across more than 650 miles (1,000 kilometers) of the Ohio River for three weeks, painting portions of the river green. Researchers have typically focused on algal blooms in lakes, but this dataset could help scientists quantify some trends in river chlorophyll concentrations.

“Our findings do not indicate if the color changes are good or bad in terms of water quality, but we showed that we can detect some trends,” said Gardner. “The next step is to investigate what humans are doing to cause those changes and whether it’s an improvement or degradation.”

For more information visit https://landsat.visibleearth.nasa.gov/view.php?id=147999.

NASA Earth Observatory images by Joshua Stevens, using data courtesy of Gardner, J., et al. (2020) and Landsat data from the U.S. Geological Survey. Story by Kasha Patel.

asprs
THE IMAGING & GEOSPATIAL INFORMATION SOCIETY

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, Remote Sensing Applications, and Unmanned Autonomous Systems Division. 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, 8550 United Plaza Boulevard, Suite 1001, Baton Rouge, LA 70809, including inquiries, memberships, subscriptions, changes in address, manuscripts for publication, advertising, back issues, and publications. The telephone number of the Society Headquarters is 301-493-0290; the fax number is 225-408-4422; web address is www.asprs.org.

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

SUBSCRIPTION. For the 2022 subscription year, ASPRS is offering two options to our *PE&RS* subscribers — an e-Subscription and the print edition. e-Subscribers can add printed copies to their subscriptions 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 print and e-Subscription with print renew on a calendar-year basis.

The rate for a Print subscription for the USA is $1105.00 USD, for Canadian* is $1164.00 USD, and for Non-USA is $1235.00 USD.

The rate for e-Subscription (digital) Site License for the USA and Non-USA is $1040.00 USD and for Canadian* is $1089.00 USD.

The rate for e-Subscription (digital) plus Print for the USA is $1405.00 USD, for Canadian* is $1464.00 USD, and for Non-USA is $1435.00 USD.

*Note: e-Subscription, Print subscription, and e-Subscription plus Print for Canada includes 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, 8550 United Plaza Boulevard, Suite 1001, Baton Rouge, LA 70809. 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*. To receive a print copy of *PE&RS* there is an additional postage fee of $60.00 USD for U.S. shipping; $65.00 USD for Canadian shipping; or $75.00 USD for international shipping per year. Annual Individual Member dues for members residing in the U.S. and Other Foreign Members are $150.00 USD and $158.00 USD for Canadians. Annual Student Member dues for members residing in the U.S. are $50.00 USD; $53.00 USD for Canadian; and $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.

GIS Tips & Tricks

Al Karlin, Ph.D., CMS-L, GISP

Disappearing Layers? – Here's a Quick Fix

For those who have been following this column, I frequently turn to my work colleagues or my students for Tips & Tricks with various GIS software packages. This month's "tip" originated with an issue encountered by several of my students. While I encourage them to use the Esri Basemaps for their work and, of course, although I advise them to use the "light gray" or other simple basemaps as backgrounds for their data, many prefer to use images as backgrounds for their maps.

After making multiple maps over a period of several weeks, the students started noticing that vector layers (feature classes or shapefiles) would disappear. When the image basemap was disabled, the vector layers would suddenly reappear. I call this the "Disappearing Layer Syndrome". After several frustrating trials, they could not make both the vectors and the basemap appear simultaneously on their maps. What could be happening?

If you have ever had this syndrome, or when your GIS software starts to run really slowly the solution is really simple. What is happening is that the computer's memory cache dedicated to software has been consumed. To repair the issue, just clear the cache manually.

For ArcGIS Desktop

To manually clear the cache in ArcGIS, Use the Customize | ArcMap Options (as in Figure 1).

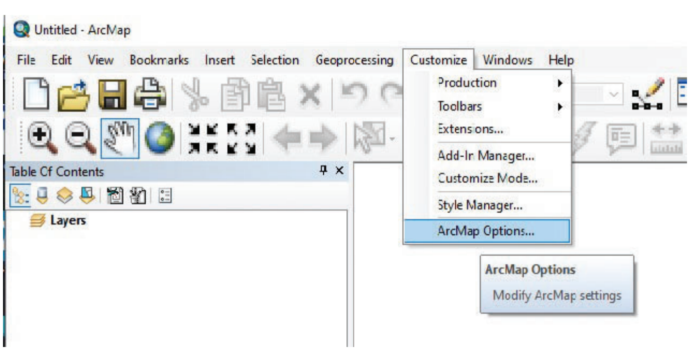


Figure 1. Selecting the ArcMap Options in ArcGIS 10.8.

And then choosing the "Display Cache" Tab from the menu options. As in Figure 2, this tab will show you how much memory is being used and by pressing the "Clear Cache" button, you will manually clear the cache. I recommend checking this area regularly, and clearing the cache when the "Currently Used" value exceeds 200 MB. Of course, you can experiment on your computer to find the optimum cache size.

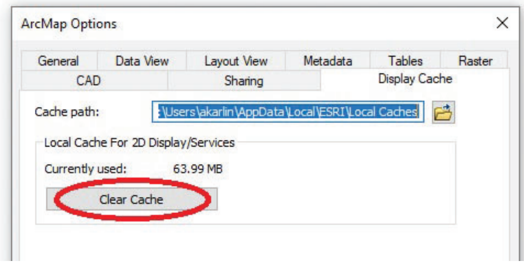


Figure 2. The Display Cache Tab on the ArcMap Options window shows that I am currently using about 64MB of memory.

For ArcGIS Pro

Use the "Project" tab and select "Options" to activate the Options menu. Then choose "Display" to show the display options. In the example below (Figure 3), my application is currently using almost 500 MB of cache (remember, this is ArcGIS Pro, a 64-bit application). By checking the box, it will clear that memory.

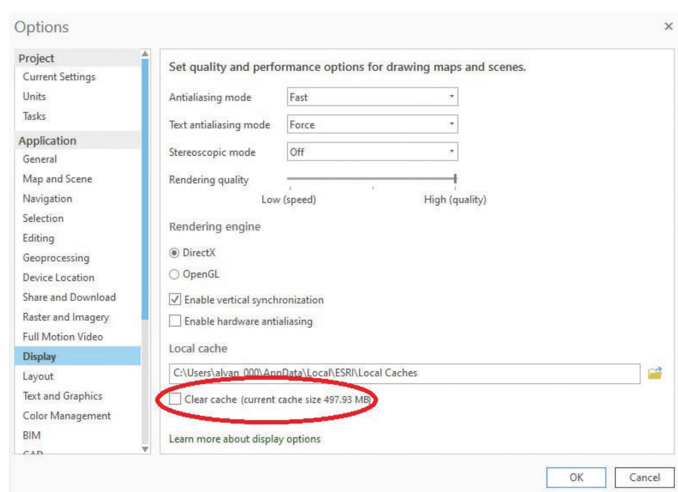


Figure 3. Using the Options | Display menu to clear the cache in ArcGIS Pro.

continued on page 13

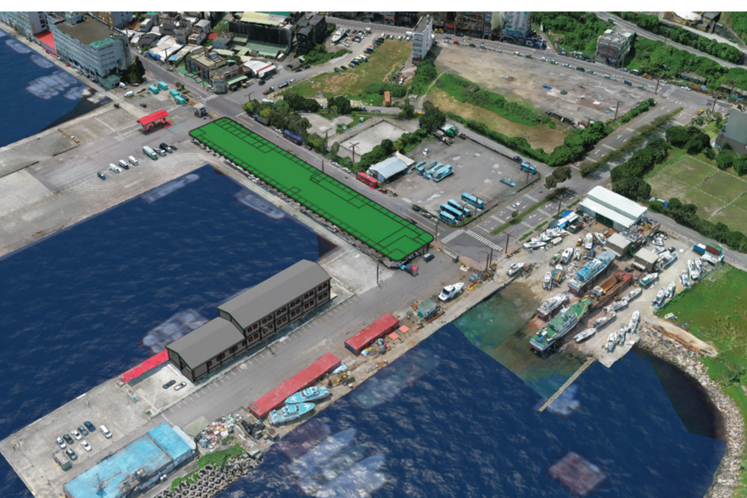
Photogrammetric Engineering & Remote Sensing
Vol. 88, No. 1, January 2022, pp. 5-13.
0099-1112/22/5-13

doi: 10.14358/PERS.88.1.5

The Arrival of Disappearance and The Map of the Future

By Lawrie Jordan
Esri Corporate Director, Imagery and Remote Sensing

Recently a long-time friend offered me a ride in his new Tesla, and as a car buff I naturally jumped at the chance. The vehicle's high-resolution map display, combined with its extensive array of advanced remote sensors and processors, elegantly simple situational awareness, feature identification, and spatial agility for autonomous operation, made me feel like I was actually sitting inside a modern "GIS on wheels." This remarkable integration of multiple technologies is giving us a front seat view into the transformation of society's future mobility, in much the same way that GPS has transformed society into now being "never lost." This experience started me wondering whether this might be a metaphor for the rapidly accelerating evolution and integration of GIS and remote sensing, and what we might see just around the corner in mapping technology.



A digital twin of a fishing port created using Site Scan for ArcGIS.

The Great Disappearing Act

Some readers will fondly remember (ok, maybe not so fondly) the early days of computing, when mainframes were incarcerated in refrigerated cells and required specialized knowledge to operate. They were entities unto themselves, monolithic in size, cumbersome to interact with through punched cards and 9-track tapes, and they catered to an esoteric group possessing rarified skills.

A major jailbreak occurred in the early 1970's when the machines escaped from their cells and began to get smaller. And more powerful. And more accessible. These 32-bit mini-computers of the 70's such as the Vax, PDP, Sun Microsystems, Prime, and others soon gave way to a new generation of personal computers in the early '80s, bringing the machines into the homes of people. Much of this new audience had little or no technical background, and had little interest in computers per se—they were just interested in what the machines could do. Moore's law accelerated, and microchips began to migrate into automobiles, appliances, thermostats, and lots of other devices

Computers were soon everywhere, and interestingly, they were no longer called "computers," as they were often hardly even noticeable. They reached maturity even as they vanished from sight. Conveniently, the massive data sets they needed to access and operate on, including imagery and maps, also "disappeared" into the cloud and became instantly available at scale. Having seen this unfold over the last few decades, I would assert that a technology has truly arrived when it actually disappears, and becomes a new normal and an expectation in everyday life. Nowhere is this more evident than in the smart phone. We all carry around these cleverly disguised super-computers with their ever-expanding library of apps, using them daily at work and at home, and few could imagine life without one today. The "computer" as we once knew it has truly arrived through its disappearance.

The Illusion of Simplicity

All of the great benefits of technology aside, a fundamental requirement for adoption of any tech at scale is that it must be easy to use by non-technical users. The GPS system mentioned earlier is a shining example of what I refer to as "The Illusion of Simplicity." Society was forever changed when several of the most complex technologies ever invented in human history converged to form the GPS-based navigation system.

Initially developed by and for the military, the combination of a GPS satellite constellation, topologically structured intelligent maps, address matching, and advanced routing algorithms all formed a new foundation for how global society navigates from place to place. This appears to the user as being elegantly simple – it knows where you are, takes you where you want to go, helps you if you get lost, gives you options for routing, and all of this in turn-by-turn instructions in a choice of different voice types. Push the "Home" button and you're never lost – it just works.

Although enormously complicated behind the curtain, thanks to this "illusion of simplicity" in interface design, both the GPS system and smart phones have been adopted globally by professionals and consumers alike. I would assert that this aspect of simplicity will be a key element going forward in the next generation of GIS adoption at scale.

The Power of Three

The early years of GIS and Remote Sensing technologies saw the two as linked but co-evolving along parallel but separate trajectories. GIS spoke the language of points, lines, and polygons, while Imagery was measured out in pixels, rasters, and point clouds. Today, modern GIS systems fully integrate Imagery and Remote Sensing capabilities at all levels.

GIS and building information modeling (BIM) software were also linked—tentatively, at first, and more directly as the years progressed. Many of the overall objectives of the technologies were, of course, quite similar. Gradually, GIS, Remote Sensing, and BIM each became more fluent in the language spoken by the other, including the recent development of the GeoBIM concept and related products. As GIS continues to evolve and integrate these essential elements, the glass walls that once separated the three are beginning to "disappear", setting the stage for a new "arrival" and an expanded definition of "what is a map," and "what I can do with it." Simply stated, these three technologies are "better together," and collectively they deliver to us a compelling Geographic Advantage.

The Map of the Future

Einstein famously once said "If I can't see it I can't understand it." We see in 3D, and this helps us to better understand everything around us. The vast majority of maps created in the early days of GIS were based at some early point in their creation on imagery of some type. Features were digitized and the image was abstracted to form a simplified 2D line drawing with just the features of interest. With the advent of stereo imagery, LiDAR sensors, drone technologies, 3D meshes, structure from motion, game engine processors, and other capabilities, GIS systems today have become 3D, and this has greatly assisted in solving certain classes of geospatial problems and design challenges that can only be understood and addressed in 3D, notably in dense urban settings.

I would submit that a new and expanded definition of a "map" would be based on a synthesis of all of the items discussed above, and the two graphic illustrations here give us a preliminary glimpse. Specifically, I would describe this "Map of the Future" as a photo-realistic, intelligent, 3D image that includes full GIS attributes, which I can interrogate, fly

3D mesh of Zurich, Switzerland creating using Site Scan for ArcGIS.

around and through, visualize in any format and dimension that makes sense to me, and have available anytime on any device. Plus, its architecture is designed to be infinitely scalable in any coordinate system, including underground and ocean floor. This leverages the power of three above (GIS, Remote Sensing, and BIM) and is wrapped in an envelope of simplicity.

Since all of the capabilities listed above to create this "Map of the Future" exist today in modern GIS systems, you might reasonably ask the question "So, what's missing?"

So What's Missing?

A new language. We need a new language to describe these maps of the future. Our time-honored traditional language of map scales (ie. 1:24,000) and resolution (ie. 30 cm. pixels) is simply inadequate to describe and accurately communicate information about a high-resolution 3D image from multiple viewing angles. For example, what is the "scale" of a perspective view? Well, you could say that it's infinitely variable based on certain parameters, but that's not really very helpful in the end. Further, I don't think that the "level-of-detail" ontology (ie. "LOD-1" for solid blocks, "LOD-2" for adding roof detail, etc.) is sufficient when we can now produce exquisite drone-based 3D meshes and "Digital Twins" in remarkable detail, as shown in these examples.

I think exploring the need for a new and expanded language in mapping could be a great area of focus that our professional organization (ASPRS) in partnership with the Academic community and industry could provide much needed leadership.

So, in brief conclusion, it's my sense that we're just at the beginning of an entirely new and bright chapter in the geospatial community, with some old barriers in the road "disappearing" and some remarkable new "arrivals" just around the corner on the road ahead. This promises to be a ride worth taking!

continued from page 2

- the Trimble post-processed CenterPoint® RTX™ correction service (PP-RTX) for centimeter-level mapping without GNSS reference stations,
- In-air development of raw imagery to JPEG-ready files for creating map products immediately upon landing, and
- Lead'Air's innovative X-Track flight management, which enables the system to be flown outside of planned flight lines to follow roads, rivers and coastlines.

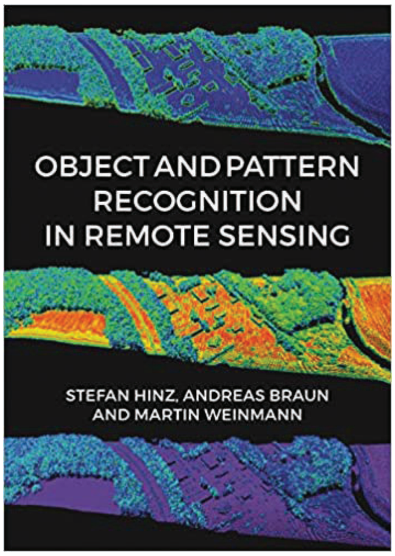
Applanix's DG technology suite provides direct GNSS inertial georeferencing, meaning that all pixels in the aerial images taken by NOAA are mapped at their exact location on the ground.

"We have worked with Applanix for nearly 20 years," said Michael L. Aslaksen Jr., chief of the remote sensing division of NOAA's National Geodetic Survey. "The level of sophistication they bring to aerial imagery and mapping keeps our team at the forefront of the industry. Their customer support team is always open to new ideas, new innovations and doing whatever it takes to get the job done."

First responders have access to this imagery and mapping within 24 hours via the cloud (as does anyone at https://storms.ngs.noaa.gov/; zoom in for the detailed images) and can map detailed response plans based on highly accurate data highlighting where the greatest need lies. Access to this turnkey emergency response imagery is available to any federal agency, municipality, insurance companies and other entities who depend on highly accurate information to plan for and recover from disasters.

CALENDAR

- 1-3 February, 2022, **URIS LEAP Conference**. For more information, visit www.urisa.org/leap.
- 6-8 February 2022, **Geo Week 2022**, Denver, Colorado. For more information, visit www.geo-week.com/.



Object and Pattern Recognition in Remote Sensing

Edited by Stefan Hinz, Andreas Braun and Martin Weinmann.

Whittles Publishing, Dunbeath, Caithness, UK. 2021. xiii and 350 pp, 88 color and 37 black and white illustrations, 18 tables, index. Hardcover. ISBN 978-1-84995-128-9. $183.96. £90.00; Amazon $107.73.

Reviewed by Stewart Walker, sole proprietor, photogrammetry4u, San Diego, California.

Many geospatial luminaries have emerged from the academic apparatus in Karlsruhe. Of the 16 contributors to *Object and Pattern Recognition in Remote Sensing* book, all but Franz Rottensteiner are educated and/or work at Karlsruhe Institute of Technology (KIT). They provide a condensed, masterly review of a considerable body of material representing many strands of remote sensing research. The subtitle is *Modelling and Monitoring Environmental [sic] and Anthropogenic Objects and Change Processes*.

Stefan Hinz, head of KIT's Institute for Photogrammetry and Remote Sensing (IPF), is well known and is president of ISPRS Technical Commission I (Sensor Systems). It's appropriate, therefore, that a foreword has been contributed by ISPRS president Christian Heipke, giving perspective, in terms of both remote sensing and photogrammetry and also the eminence of KIT.

The book is in three parts. The first, "Methodology," begins with an introduction by Hinz, setting the stage. He proceeds with Chapter 2, "Object Data and Sensor Modelling" a readable synthesis of vast amounts of material, and includes a large number of references at the end of the chapter - this happens with every chapter and immeasurably increases the book's value. This excellent review material continues with Chapter 3, by Martin Weinmann, "Feature Extraction from Images and Point Clouds: Fundamentals, Advances and Trends." Andreas Braun joins Weinmann for Chapter 4, "A Short Survey on Supervised Classification in Remote Sensing." Rottensteiner takes the helm for Chapter 5, "Context-based Classification" and Uwe Weidner completes the section with Chapter 6, "Toward a Framework for Quality Assessment in Remote Sensing Applications." These syntheses, complemented by the ample references, are invaluable and justify buying the book - yet little of the work cited dates from later than 2010.

Part II, "Applications," summarizes research done in Karlsruhe. Chapter 7, "From Raw 3D Point Clouds to Semantic Objects" (Weinmann, Sven Wursthorn, Boris Jutzi), focuses on terrestrial laser scanning and range cameras. The coverage of point cloud matching and registration is very useful, as is the material on feature extraction and scene interpretation. The references again are not new, however, and the datasets were last accessed in 2013. Hinz returns to the stage, with Jens Leitloff, with Chapter 8, "Traffic Extraction and Characterization from Optical Remote Sensing Data," full of well explained, interesting work based on still images rather than video. The tell-tale is a footnote, "Updated and revised version of (Hinz et al. 2008)": the material is mature and the update refers mainly to work published 2009-14. Chapter 9, "Object Extraction in Image Sequences" by Florian Schmidt and Hinz, summarizes strong work on the detection of persons from aerial photography with a frequency of 2 Hz, though this is hardly representative of fast-cycling modern cameras or the role of UAVs. Yet, like much of the material in this book, the cohesive, lucid presentation, drawing on extensive literature, provides understanding and background. Chapter 10, "A Process-based Model Approach to Predict Future Land-Use Changes and Link Biodiversity with Soil Erosion in Chile," by Andreas Braun and Callum Banfield, is based on the first author's PhD thesis and the second author's MS thesis, both at KIT in 2013. This shorter chapter is practical and gives a useful account of the challenges and solutions within a particular research theme.

The book then change sensors. Chapter 11, "Interferometric

Photogrammetric Engineering & Remote Sensing
Vol. 88, No. 1, January 2022, pp. 9-10.
0099-1112/22/9-10

doi: 10.14358/PERS.88.1.9

SAR Image Analysis for 3D Building Reconstruction," by Antje Thiele, Clémence Dubois and Hinz, takes up from Hinz's initial comments on SAR in Chapter 2 and quickly dives deep with descriptions of how buildings can be extracted from the data. Chapter 12, "Detection and Classification of Collapsed Buildings after a Strong Earthquake by Means of Laser Scanning and Image Analysis," by Miriam Hommel and Thomas Vögtle probes the practicalities of assessing damage from point clouds and imagery. This is a fascinating read and one is conscious of the human tragedies behind the science. Ulrike Sturm-Hentschel, Braun and Hinz end the section with Chapter 13, "A Settlement Process Analysis in Coastal Benin: Confronting Scare Data Availability in Developing Countries," reporting high-quality research work, using, for example, QuickBird data. The authors' complaints about lack of data, however, are less worrisome in 2021, since the constellations of multiple satellite operators provide a plethora of information with shorter and shorter repeat times.

The book is drawn together in Part III, "Conclusion." This material is more up to date than Part II. Chapter 14, "Benchmarking: a basic requirement for effective performance evaluation," by Weinmann and Braun, stresses the importance of standard data sets for assessing new approaches and gives several examples. Seven authors worked on Chapter 15, "Remote sensing and computer vision image analysis: summary and recent trends." This title confirms that the emphasis is less on photogrammetry than some readers would perhaps prefer. Sometimes whimsical, with glimpses of humor, this chapter weaves the book's threads into fabric. It levers the book into the second half of the 2010s and is a perceptive assessment of trends. The authors' experiences and involvements shine through and there is acuity as well as description.

Your reviewer's unease with the currency of the material, nevertheless, resurfaced in the final chapter. There is excellent but brief coverage of deep learning, for example, which gave your reviewer more insight than many of the heavy papers on the subject with which he has grappled. Yet between this (pages 338-339) and the comments on page vii in Heipke's foreword and pages 6-7 in Hinz's introduction, the topic is barely mentioned. How I wish there had been much more on this topic! There is an insight on page 339 that the authors "finalized the book in 2018/19," which suggests that the final steps to publication were lengthy ones. There is a remark on page 336 about "recent reviews" of multiple classifier systems that were published in 2002 and 2007.

Object and Pattern Recognition in Remote Sensing is a fine, well produced book, a real pleasure to use. The Scottish firm, Whittles Publishing, has incorporated both monochrome and color graphics that are attractive, though sometimes on the small side (the legend of figure 10.3 has a 3-point font!). There are few typos, though tighter editing would have eliminated some minor curiosities in language. Though rather advanced for students new to remote sensing, it will certainly serve the lecturers, practitioners, researchers, advanced undergraduates, and postgraduates that the publisher's blurb on the back cover suggests are the target market. While the absence of material on the last ten years, particularly in Part II, must remain a demerit until the second edition, the excellence of Parts I and III should convince doubters to purchase this book.

GRIDS & DATUMS

REPUBLIC OF VANUATU

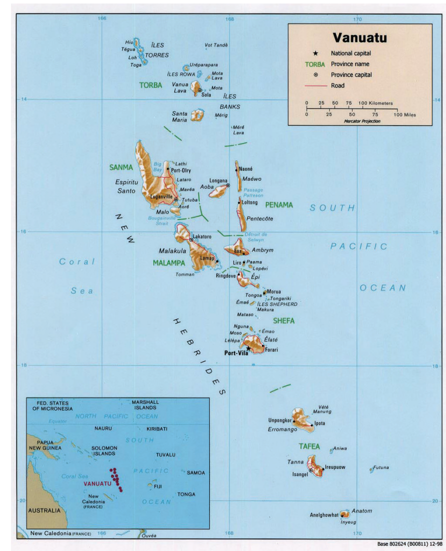
BY Clifford J. Mugnier, CP, CMS, FASPRS

The Grids & Datums column has completed an exploration of every country on the Earth. For those who did not get to enjoy this world tour the first time, *PE&RS* is reprinting prior articles from the column. This month's article on the Republic of Vanuatu was originally printed in 2004 but contains updates to their coordinate system since then.

Inhabited for thousands of years by Melanesians before discovery by the Portuguese navigator Pedro Fernandes de Queirós, the islands were forgotten for 160 years and were then visited by the French navigator Louis-Antoine de Bougainville in 1768. The English mariner Captain James Cook explored the islands in 1774 and named it the New Hebrides. "The British and French, who settled the New Hebrides in the 19th century, agreed in 1906 to an Anglo-French Condominium, which administered the islands until independence in 1980." What the World Factbook doesn't say is that the local people referred to it as the Pandemonium! Vanuatu \ vän-, -wä-tü \, is a group of more than 80 islands in the southwest Pacific Ocean northeast of New Caledonia and west of Fiji (*PE&RS,* October 2000). With a land area of 12,200 km^2, the republic is slightly larger than Connecticut.

Vanuatu has a tropical climate, the terrain is comprised mostly of volcanic mountains with narrow coastal plains, the lowest point is the Pacific Ocean, and the highest point is Tabwemasana (1,877 m) on the island of Espiritu Santo. The total coastline is 2,528 km and its maritime claims are based (naturally) on archipelagic baselines. The exclusive economic zone is 200 nautical miles (NM), the territorial sea is 12 NM, the contiguous zone is 24 NM, and the continental shelf claim is 200 NM or to the edge of the continental margin – all of these claims are customary and are recognized under the International Law of the Sea.

In the Vanuatu Geodetic Control Network Report by Bakeeliu, Kanas, and Kalsale in June 2001, the network that began in the 1960s was generally detailed to the pres-



ent. The Institute Géographique National (IGN) of France started their network in the 1960s. "The IGN network was made in two blocks, one of which covers the islands of Santo, Aoba, Pentecost, Maewo, Ambrym, Malekula Epi, Éfaté in the northern part of Vanuatu while the other block covers Erromango, Tanna, Anatom and the nearby small islands in the south. The islands left out were the Banks and Torres group in the far north of Vanuatu." The report continues, "The IGN [*datum – Ed.*] was based on the astronomical observation made at Bellevue on Éfaté." (Note that another common spelling for the island of Éfaté is Île Vaté). The Vanuatu (IGN) 1960 Datum origin coordinates at Bellevue are Φ_o = 17°44´17.40˝ South, Λ_o = 168°20´33.25˝ East of Greenwich, and the ellipsoid of reference is the International 1909 (Madrid 1924) where a = 6,378,388 m, and $^1/_f$ = 297. The National Geospatial Intelligence Agency (NGA) lists the transformation parameters from The Vanuatu (IGN) 1960 Datum (Bellevue) to the WGS84 Datum as Δa = –251m, Δf = –0.14192702, ΔX = –127m±20m, ΔY = 769m ±20m, and ΔZ = +472m±20m. This relation is based on observations at three stations. John W. Hager, retired from what is now NGA says, "The transformation states that

Photogrammetric Engineering & Remote Sensing
Vol. 88, No. 1, January 2022, pp. 11-13.
0099-1112/22/11-13

doi: 10.14358/PERS.88.1.11

it is for the islands of Éfaté and Erromango but my notes imply that the trig list applies from Éfaté to Espiritu Santo. Source for this is 'Loose Minute, MCE Ht/KHG/PL, 10 July 1970 with one page from I.G.N. Trig list.' I don't remember what all the letters mean but it was a letter from the British Military to U.S. Army Topographic Command." Hager also commented that, "Pier Observation Spot, Vila Harbor, Éfaté Island. I have no data but questioned for my further investigation whether is was Fila or Vila, Fila possibly being a corruption of Éfila." Presumed datum point is S. T. 1 (Service Topographique) at latitude 15°17´16˝S., longitude 16758´34˝E. This was taken from 'Traverse Around Perimeter of Aoba Island,' 30 Sept. 1969." He found a 1949 IGN reference to a local grid for the Aoba Island Datum where False Northing = 12,000 meters and False Easting = 18,000 meters. According to Hager, "the only odd map projection I find is for Nouvelles Hébrides, Fuseau Calédonie-Hébrides, Gauss projection, [*Transverse Mercator – Ed.*], International ellipsoid, meter, latitude of origin = equator, longitude of origin = 167°E, scale factor unknown but probably unity, false northing (y) = 2,600,000 meters, false easting (x) = 1,000,000 meters. This is from 'Catalogue de Cartes en Service Publiées par l'Institute Géographique National,' Paris, 1 July 1949."

Referring back to the Vanuatu Geodetic Control Network Report, "The adjustment used by DOS [*Directorate of Overseas Surveys, UK – Ed.*] was initiated from the same points as the IGN however the astronomical observations and adjustment was done separately. The DOS adjustment covers the islands of Santo, Aoba, Maewo, Pentecost, Ambrym, Malekula and Pamma in the north and Éfaté, Erromango, Tanna, Anatom and Futuna in the south. The DOS however extended its triangulation further throughout the country covering and strengthening the network to other islands, except Bank and Tores in the far north. This adjustment was used for mapping as well as cadastral. DOS adjustment uses the same scale factor of one (1.00000) throughout the country, though each island has its own origin."

Continuing, early in the "1980s the Vanuatu Government attempted to connect the DOS north and south block using traverse methods with the introduction of Telurometer distance measurements. However, it was found that there was some discrepancy between the two blocks. It was uncertain then that the error was in the traverse observation or the astronomical observation of the two blocks. It was also difficult to undertake alternative method of triangulation as the sights between Epi and Emae islands was very difficult. It was seen that it may be easier if a triangulation was done through the islands between the two blocks, however for some reason this was not done. The technology at that time may also be the cause of the inaccuracy of the observations. In mid 1990s the Australian Government assisted the Vanuatu Government by providing funds, technology and human resources through the Australian Defense Cooperation to run a Doppler network that covers the whole country. This has enabled the Vanuatu Government to anticipate the strengthening of the country's survey control network on the WGS72 spheroid. The network was produced to control the aerial photography of the country. For cadastral purposes the DOS geodetic adjustment is still maintained."

I asked Russell Fox, now retired from the International Geodetic Library of the Ordnance Survey International, United Kingdom, if he had anything to help me on my column on Vanuatu. To my (usual) surprise, he certainly did have something. Fox had worked there for three years! "The Condominium (known as the Pandemonium locally) was a strange form of government, the British and French running parallel but separate administrations in the same territory (so not analogous with St. Maarten/St. Martin). There were French and British police forces, hospitals, schools, etc. Residents had to use "their" facilities. Citizens of countries other than Britain (& Commonwealth) or France had to opt for either honorary British or honorary French status and use the appropriate services. This split the local people also, half of whom were educated in the French milieu and half in British traditions. There was "trouble in paradise," as the newspapers put it, during the immediate pre-and post-independence period, as the more radical and pro-independence English-speaking ni-Vanuatu jostled for power with the French speakers and French settlers, who preferred the status quo (not least because French plantation owners would be most affected by proposed changes in land tenure)."

Fox continued, "I worked in Vanuatu from 1983-86. Independence had come in 1980, so I did not personally witness this, but one of the Survey Department's tasks pre-independence was to measure the heights of the flagstaffs at the British and French Residencies in Port Vila. There would have been a diplomatic incident if either the Union Jack or the Tricoleur had been flown slightly higher than the other! The Condominium was the result of Anglo-French rivalry in the Pacific during the late 19th century; I believe that the Australian colonies were particularly keen to avoid a French takeover of the New Hebs as well as New Caledonia and they lobbied the British Govt. to do something about it. The answer was Condominium, if only to avoid an Anglo-French war. Another Condominium was the Anglo-Egyptian Sudan. The WWII US presence in the New Hebs was still evident in the 1980s, with 6-wheel trucks on plantations, USN dustbins [galvanized trash cans?] being used as water containers and metal plates from airfield runways being used as property fences." [*I remember seeing the same things when I lived in Panamá – Ed.*]

"The main post-1978 survey activities I know of were: 1980 – A dozen Doppler stations were observed by 512 Specialist

Team, Royal Engineers. 1983-86 'Operation Algum' – major support for the Survey Department was received from the Royal Australian Survey Corps. This involved a Doppler campaign throughout the islands, new aerial photography, readjustment of the DOS and IGN trig networks on WGS84 and setting up a map production facility in the Survey Department. 1980s-1990s New Editions of the DOS 1:1,000,000 map were produced by the Survey Department, also a new 1:50,000 series. The Vanuatu Map Grid was introduced, a national TM projection to replace the assorted island grids that existed previously. The Survey Department produced a brief paper in about 1976/77 that discussed the significant differences between DOS and IGN positions in the New Hebs (nearly a km in the northern islands if I recall correctly). Those discrepancies weren't solved – or circumvented – until OP Algum, but the Survey Department did develop a TM grid (called Éfaté TM 77) for the main island, Éfaté or (Vaté), in 1977 to improve the control situation there by unifying disparate surveys and replacing the old Cassini grid. Both DOS and IGN used International Spheroid, but had datums in different places, and trig block boundaries in different places – the DOS North Block was islands North of Éfaté, and South Block was Éfaté and islands south. IGN had a North Block (Éfaté and islands North) and South Block (Erromango to Aneityum). I think the most northerly island in the New Hebs, the Banks and Torres Islands, were not reached by either the DOS or IGN networks and had local astro fixes only."

The National Geospatial Intelligence Agency (NGA) lists the transformation parameters from the Santo (DOS)1965 Datum (Espiritu Santo Island) to the WGS84 Datum as: $\Delta a = -251$m, $\Delta f = -0.14192702$, $\Delta X = +170$m ± 25m, $\Delta Y = +42$m ± 25m, and $\Delta Z = +84$m ± 25m. This relation is based on observations at one station. Thanks to John W. Hager; Russell Fox; Tony Kanas, surveyor; and, the Vanuatu Department of Land Surveys for their generous assistance.

Vanuatu Update

In 2014, the U.S. Department of State published No. 137 Limits in the Seas, *Vanuatu: Archipelagic and other Maritime Claims and Boundaries*. Coordinates are shown to four decimal places of arc seconds, and all connecting lines are defined as geodesics. No datum, ellipsoid, nor International Terrestrial Reference Frame date is stated.

Vanuatu Geodetic Control Network Report, Mike Bakeoliu, Tony Kanas, Moses Kalsale, 09 June 2001.

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

GIS Tips & Tricks, *continued from page 5*

In GQIS and other GIS software—QGIS and several other GIS software packages allow you to clear the cache through Python or directly through the command line. For QGIS, the command would look like:

```
rm -rf ~/.qgis2/cache/data7
```

Here is a link for additional help for QGIS:

https://gis.stackexchange.com/questions/356704/how-to-clear-the-cache-of-qgis-3-10-with-python

It is that easy to solve the "Disappearing Layer Syndrome".

Send your questions, comments, and tips to GISTT@ASPRS.org.

Al Karlin, Ph.D., CMS-L, GISP is 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.

STAND OUT FROM THE REST

EARN ASPRS CERTIFICATION

ASPRS congratulates these recently Certified and Re-certified individuals:

CERTIFIED PHOTOGRAMMETRIST

Scott Merritt, Certification #1662CP

Effective September 9, 2021, expires September 9, 2026

Derek Smith, Certification #1661CP

Effective September 9, 2021, expires September 9, 2026

Jamie Gillis, Certification #1660CP

Effective August 25, 2021, expires August 25, 2026

RECERTIFIED PHOTOGRAMMETRIST

Bill Swope, Certification #R1604CP

Effective September 12, 2021, expires September 12, 2026

John Matthews, Certification #R804CP

Effective May 8, 2021, expires May 8, 2026

Philip Gershkovich, Certification #R1500CP

Effective July 30, 2021, expires July 30, 2026

Sean Bolender, Certification #R1508CP

Effective October 18, 2021 expires October 18, 2026

Adam Dolberry, Certification #R1601CP

Effective August 22, 2021, expires August 22, 2026

Brandon Rench, Certification #R1603CP

Effective September 12, 2021, expires September 12, 2026

Nora Csanyi May, Certification #R1497CP

Effective July 14, 2021, expires July 14, 2026

Miles Strain, Certification #R1120CP

Effective September 21, 2021, expires September 21, 2026

RECERTIFIED MAPPING SCIENTIST – REMOTE SENSING

Michael Tuffly, Certification #R196RS

Effective September 4, 2021, expires September 4, 2026

Dr. D. Marius Necsoiu, Certification #R153RS

Effective June 13, 2021, expires June 13, 2026

CERTIFIED MAPPING SCIENTISTS LIDAR

Derek Smith, Certification #L068

Effective September 22, 2021, expires September 22, 2026

Adam Campbell, Certification #L006

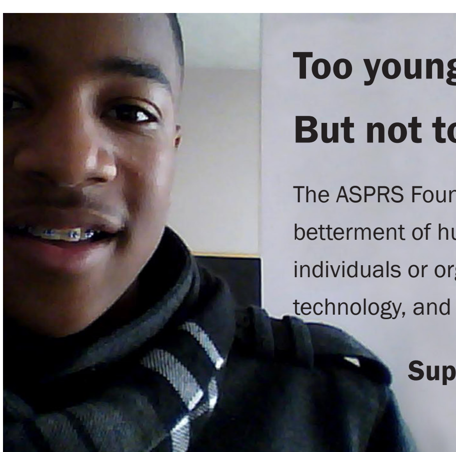
Effective March 28, 2021, expires March 28, 2026

RECERTIFIED MAPPING SCIENTISTS GIS/LIS

Levi Ladd, Certification #R282GS

Effective March 5, 2020, expires March 5, 2025

ASPRS Certification validates your professional practice and experience. It differentiates you from others in the profession. For more information on the ASPRS Certification program: contact **certification@asprs.org,** visit **https://www.asprs.org/general/asprs-certification-program.html.**



NEW ASPRS MEMBERS

ASPRS would like to welcome the following new members!

Oluwatola Ibukun Adedeji
Javier Aguilar
Mohsen Ahmadkhani
Abidemi Ayodeji Aina
Samuel Akinbola Akinnusi
Ismail Adewale Alatise
Barry Anderson
Alaina A. Baker
Cailey Bechtol
Bryce Everett Berrett
Carson Curry
Andy Osagie Egogo-Stanley
Caner Ferhatoglu
Steven Filippelli
Andres Esteban Garzon-Oechsle
Iyanu Joshua Gbiri
Craig Gooding
Christina A. Greco
Adam Gribben
Westin Guthrie
Michael Carl Haas, PhD
Andrea Renee Anderson Hurzeler
Jason Ingram
Jeffrey R. Irwin
Tyler Kaebisch
Jochen Keil
Richard Kleinmann, PLS
Eric Levenson
Qian Liu
Yajie Liu
Junaid Mushtaq Lone, PhD
Fangzheng Lyu
Marty Marquis
Michael McLane
Blair A. Mirka
Michael O'Brien
Chukwuma John Okolie
Hamed Olabode Olanrewaju
James John Patterson
Uriel Pérez Gómez
Vinicius Perin
Vikram Potdar
Timothy Prestby
Donald Joe Province
Christopher W. Reichert
Alison Ritz
Jeronimo Roldan
Dan Rose
Samuel Rose
Eliceo Ruiz Guzman
Kevin Wesley Samples
Peter L. Samson
Tracy Shane
Chelsea Sidenblad
Danielle Smilovsky
Kevin Spruill
Ariel Muir Staveley Parker
Jordan Lee Thomas
Amber Tomaras
Jennifer Towns
James Wilson
Anna Windle
Chao Xu
Caiwang Zheng

FOR MORE INFORMATION ON ASPRS MEMBERSHIP, VISIT
HTTP://WWW.ASPRS.ORG/JOIN-NOW

Improving Urban Land Cover Mapping with the Fusion of Optical and SAR Data Based on Feature Selection Strategy

Qing Ding, Zhenfeng Shao, Xiao Huang, Orhan Altan, and Yewen Fan

Abstract

Taking the Futian District as the research area, this study proposed an effective urban land cover mapping framework fusing optical and SAR data. To simplify the model complexity and improve the mapping results, various feature selection methods were compared and evaluated. The results showed that feature selection can eliminate irrelevant features, increase the mean correlation between features slightly, and improve the classification accuracy and computational efficiency significantly. The recursive feature elimination-support vector machine (RFE-SVM) model obtained the best results, with an overall accuracy of 89.17% and a kappa coefficient of 0.8695, respectively. In addition, this study proved that the fusion of optical and SAR data can effectively improve mapping and reduce the confusion between different land covers. The novelty of this study is with the insight into the merits of multi-source data fusion and feature selection in the land cover mapping process over complex urban environments, and to evaluate the performance differences between different feature selection methods.

Introduction

In recent years, with the rapid development of the economy and the improvement of urbanization, the large area of hardened land has been squeezing the urban ecological space. The urban heat island effect has become increasingly prominent, and problems such as air pollution and environmental damage have become increasingly serious threats to our living environment (Kuang *et al.* 2015; Li *et al.* 2011; Shao *et al.* 2020a). High-precision urban land cover (ULC) data is an important foundation for rational development and dynamic monitoring of land resources (Chen *et al.* 2021a). It also plays a key role in climate assessment, temperature change, environmental protection and other research, and provides scientific basis for urban planning, management and sustainable development (Huang and Wang 2020; Lazzarini *et al.* 2015; Li *et al.* 2017; Zhang and Sun 2019).

Remote sensing data, with the advantages of low cost and high efficiency, has become the main data source for land cover mapping (Friedl *et al.* 2002; Gallego 2004; Griffiths *et al.* 2019; Khatami *et al.* 2016). Basic research usually only uses optical remote sensing data to distinguish land covers according to the spectral feature differences between different classes (Shih *et al.* 2019). However, due to the complex distribution of land covers and highly mixed spatial pattern in urban areas, there are many phenomena of "different body with same spectrum" or "same body with different spectrum". Therefore, ULC mapping based only on spectral features of optical remote sensing data cannot completely identify the effective information of targets, and the accuracy of mapping results is difficult to guarantee (Hodgson *et al.* 2003; Weng *et al.* 2009). In addition, the quality of optical remote sensing images is easily affected by meteorological conditions, and the images obtained during rainy or cloudy weather are not suitable for land cover mapping. Fusing multi-source remote sensing data and giving full play to the advantage of different data is an effective way to further improve ULC classification accuracy (Ty *et al.* 2016; Prins and Niekerk 2020). Synthetic aperture radar (SAR) can observe the geometric and dielectric properties of the Earth's surface through clouds, fog and haze. Previous studies have shown that the fusion of optical remote sensing data and SAR data can realize information complementation, thus reducing the confusion between different land covers and improving the ULC classification accuracy (Joshi *et al.* 2016; Shao *et al.* 2016; Shao *et al.* 2020b; Symeonakis *et al.* 2018; Tabib Mahmoudi *et al.* 2019; Zhang and Xu 2018; Zhang et al 2018). However, the fusion of multi-source data will also lead to the increase of input feature dimension, the increase of noise, and calculation amount in the classification model, which will result in the decline of the stability and interpretability of the model (Georganos *et al.* 2018). Hence, how to obtain a concise subset from multi-dimensional data that can balance classification accuracy and model interpretability is a crucial issue.

As a part of data mining, feature selection aims at selecting subsets according to the importance of each feature to reduce complexity while maintaining or improving the performance of the model (Cai *et al.* 2018; Guyon and Elisseeff 2003). In the research of land cover mapping, feature selection methods based on filtering, wrapping and tree model all have been applied (Pal 2005; Sesnie *et al.* 2008; Zhang and Yang 2020; Zhou *et al.* 2018). However, due to the different standards of the feature selection methods, the mapping results will be different. There is no consensus on the preference of feature selection methods in land cover mapping research, and the effects of different feature selection methods still need to be compared.

In this study, the Futian District of Shenzhen City was selected as the study area. Considering the complexity of land

Qing Ding, Zhenfeng Shao, and Yewen Fan are with State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China (Corresponding author: shaozhenfeng@whu.edu.cn).

Xiao Huang is with the Department of Geosciences, University of Arkansas, Fayetteville, AR 72701.

Orhan Altan is with the Department of Geomatics Engineering, Istanbul Technical University, Istanbul 36626, Turkey.

Contributed by Alper Yilmaz, June 1, 2021 (sent for review July 14, 2021; reviewed by John van Genderen, Hongsheng Zhang).

Photogrammetric Engineering & Remote Sensing
Vol. 88, No. 1, January 2022, pp. 17–28.
0099-1112/22/17–28

doi: 10.14358/PERS.21-00030R2

covers distribution in urban areas, Sentinel-2 multi-spectral data and Sentinel-1 SAR data were selected for ULC mapping. The fusion of optical and SAR data can provide effective information for land cover mapping from different perspectives. In terms of feature extraction, the original band features, index features and texture features of optical and SAR data were used to fully reflect the difference between different land covers. At the same time, in order to guarantee the classification accuracy and reduce the complexity of input features, removing features with low standard deviation (RFLSD), Chi2, ReliefF (Robnik-Sikonja and Kononenko 2003), recursive feature elimination (RFE) (Liu and An 2020), random forest (RF) (Breiman 2001) and Extra tree (Geurts *et al.* 2006) were respectively adopted to realize the feature subset selection. Then, this study compared and evaluated the performance of different feature selection methods from a series of indicators such as the overall accuracy, producer's accuracy, user's accuracy, and kappa coefficient of land cover mapping results.

Traditional classification models, such as maximum likelihood estimation and k-nearest neighbor, are mostly based on pattern recognition technology. The disadvantage of these models is that they need a large number of training samples. However, in the research of land cover mapping, due to the limitation of samples size, these traditional models have difficulty achieving ideal results. In this study, the support vector machine (SVM) model based on statistical learning theory was selected to map land covers in Futian District (Cortes 1995). The SVM model uses kernel function to transform input features, and uses support vectors to delineate hyperplane, thus maximizing the distance between classes. As a supervised classification method, it has been widely used in land cover mapping research (Clinton *et al.* 2015; Foody and Mathur 2006; Huang *et al.* 2002; Sukawattanavijit *et al.* 2017).

Because of the spectral mixture and complex spatial pattern, it is very difficult to map the land covers in urban areas. To solve this problem, this study proposed a multi-source remote sensing data fused ULC mapping framework based on hyperparameter optimization SVM model. By optimizing feature subset, the complexity of the model was reduced, and the accuracy of the mapping results was improved. The novelty of this study is with the insight into the merits of feature selection in the land cover mapping process over complex urban environments, and a comparison of the performance differences of different feature selection methods. In addition, the effectiveness of multi-source remote sensing data fusion in land cover mapping was verified. Finally, the high-precision ULC mapping results in Futian District of Shenzhen were obtained.

Study Area and Data

The Study Area

Shenzhen City, located in the south of Guangdong Province, China, is a national economic center and international city. As a special economic zone, Shenzhen's land covers have undergone profound and irreversible changes under the influence of large-scale industrialization and urbanization. The large-scale expansion of construction land on the city has led to a series of problems, such as rising temperature, falling humidity, and visibility. Land covers data is not only helpful to urban planning and development, but also very important for studying the impact of urbanization on the ecosystem. However, the rapid development of Shenzhen has made it challenging to accurately map land covers.

In this paper, Futian District, the downtown area of Shenzhen, was taken as the study area. Futian District is located in the south-central part of Shenzhen, facing Hong Kong across the river, with various classes of land cover. The geographical location of study area are 22°30′N to 22°36′N and 113°59′E to 114°06′E (Figure 1). The total area of study area is about 78.8 square kilometers. The topography in the study area is high in the north and low in the south, and the landform type is mainly hilly. In addition, the study area has a subtropical maritime climate, with the average annual temperature of 24°C, the highest temperature of 36.6°C, and the lowest temperature of 1.4°C. The average annual rainfall is 1948 mm, and the rainfall from April to September accounts for 84% of the total annual rainfall.

Data Preparation and Preprocessing

In order to accurately map the land covers in Futian District, Sentinel-2 (optical) and Sentinel-1 (SAR) images from the Sentinels Scientific Data Hub with imaging time of October 2020 were used in this study. Sentinel-2 images are bottom-of-atmosphere reflectance data after radiometric calibration and atmospheric correction (Level-2A). The corresponding

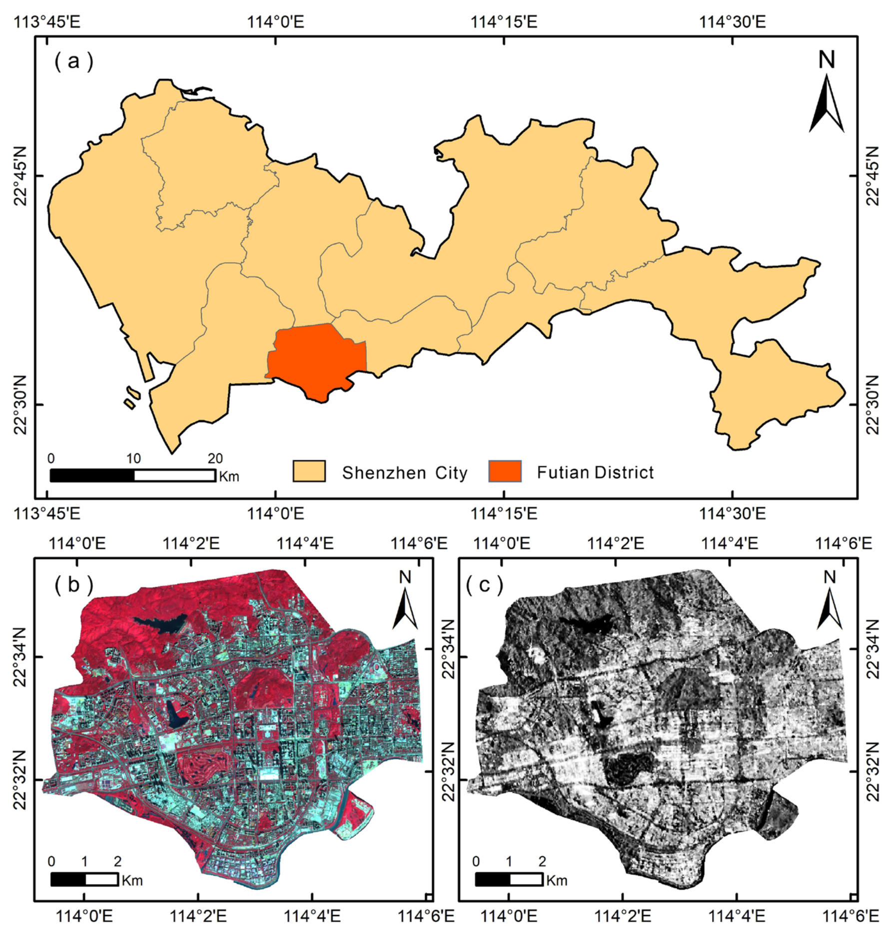


Figure 1. (a) The geographic location of study area; (b) standard false color composite (Red: band 8, Green: band 4, Blue: band 3) Sentinel-2 image in the study area; (c) backscattering coefficient under vertical-vertical polarization mode of Sentinel-1 image in the study area.

Sentinel-1 images are vertical-vertical (VV) and vertical-horizontal (VH) polarized in the interferometric wide swath (IW) mode. The wavelength of Sentinel-1 images is about 5.6 cm. After preprocessing such as multi-looking, filtering, geocoding, and radiometric calibration, the backscattering coefficients under VV and VH polarization modes were obtained. Then, the spatial resolution of both Sentinel-1 and Sentinel-2 data was resampled to 10m.

In addition, pixel level registration of Sentinel-1 and Sentinel-2 images was achieved based on polynomial model by manually selecting control points (Figure 1). Both optical and SAR images were co-registered to the same reference system of Universal Transverse Mercator projection (Zone 49N) with the datum of world geodetic system 84.

Considering the ground conditions and the resolution of remote sensing data, the study area was divided into six classes of land cover, namely, impervious surface, bare land, grass, forest, water and shadow area. According to the spectral variability, shape, area, spatial distribution, and other attributes of various land covers, a total of 92 regions of interest (ROI) were selected in the study area, and 7918 representative pixels were used for SVM training (see Table 1). In the process of selecting training samples, pure pixels were chosen as much as possible. At the same time, the training samples fully represented the spectral domain of various land covers.

Table 1. The land cover classification scheme and the size of training samples.

Class	Description	ROI	Pixel
Impervious surface	Buildings, roads, square, industrial areas	28	3250
Bare land	Exposed rock and soils	11	668
Grass	Parks, lawns and golf courses	14	673
Forest	Shrub, broadleaf and coniferous	8	875
Water	Rivers, reservoirs, ponds	9	1122
Shadow area	Areas shaded from light by tall buildings	22	1330

Notably, the impervious surface in urban areas is widely distributed and rich in types, so more training samples are needed to characterize it. To avoid the decline of model accuracy caused by too many samples of a single type, the impervious surface was divided into two categories: building and hardened ground, which were extracted separately and merged in the final mapping results. In addition, the characteristics of bare land, grass, and forest in the urban areas are obvious, so we chose fewer samples to reduce the calculation of model training. Because the waters appear in two forms: bright (rivers) and dark (reservoirs and ponds), and the shadow areas with complex shapes are mostly located near buildings or roads, the training samples were slightly increased to fully reflect their characteristics. Hyperparameter optimization of the SVM model to select the optimal penalty coefficient and kernel coefficient can effectively suppress the slight imbalance of training samples. Through the above sample selection strategy, it can not only fully characterize different land covers, but also help to improve the model efficiency and classification accuracy.

Methodology

In order to improve ULC mapping, this study extracted the band features, index features, and texture features from Sentinel-1 and Sentinel-2 data. Ideally, more features could be used to accomplish tasks better, but the reality is that appropriate features have better results (Chen *et al.* 2021b). The multi-feature fusion of multi-source data will lead to the increase of data dimension, which may cause problems such as the noise increase, existence of negative features, and difficulty of calculation convergence(Simonetti *et al.* 2014; Zhu *et al.* 2012) . This study evaluated six feature selection methods and obtained different feature subsets according to the importance of features.

Based on the hyperparameter optimization SVM model, land cover mapping results in the study area were obtained by using different feature subsets. Then, the accuracy of mapping results was assessed to verify the effectiveness of feature selection. The performance differences of six feature selection methods were also evaluated. Finally, compared with the mapping results of single source data, the advantages of multi-source data fusion for land cover mapping were verified. Figure 2 shows the overall workflow for ULC mapping based on multi-source remote sensing data fusion and feature selection.

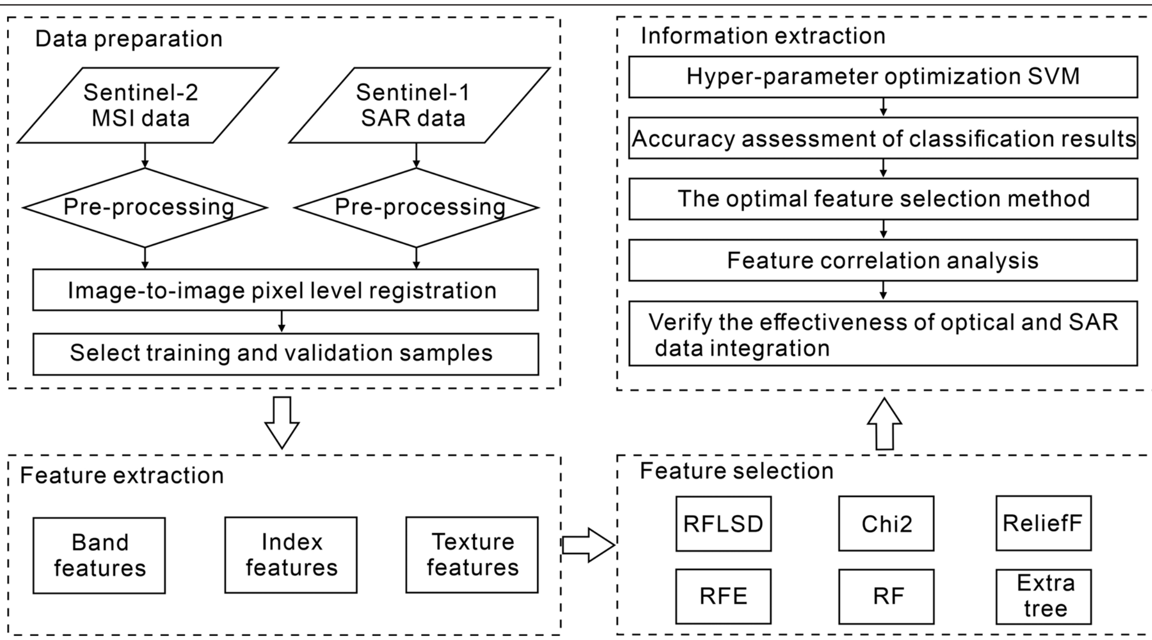


Figure 2. The overall workflow for urban land cover mapping based on multi-source data fusion and feature selection.

Feature Extraction

For Sentinel-2 data, besides the original band features, this study also extracted some spectral index features which are widely used and have good indication, such as normalized difference vegetation index (NDVI), normalized difference water index (NDWI), etc. In addition, texture features can reflect the homogeneity of images, which also play an important role in ULC mapping. First, principal component analysis on the band features of Sentinel-2 data was conducted. Second, the gray-level cooccurrence texture method was used to extract texture features from the first component.

Similar to optical data, in addition to backscattering coefficients, the index features and texture features of Sentinel-1 data were also extracted for ULC mapping. A total of 46 features were extracted from optical and SAR data, and the values of each feature were converted to values between 0 and 1. The detailed description is shown in Table 2.

Feature Selection

Removing Features with Low Standard Deviation

Standard deviation can reflect the dispersion degree of a group of data. When the standard deviation is small, the data within the group contains less information, which is closer to the average. In mapping research, features with low standard deviation cannot reflect the differences between different classes well. The standard deviation of feature t is shown in Equation 1:

$$S_t = \sqrt{\frac{\sum_{i=1}^{N}(t_i - \bar{t})^2}{N-1}} \tag{1}$$

where N is the number of training samples, t_i is the feature t in the i-th sample, and $\bar{t}$ is the average of feature t in all training samples.

Chi2 Test

Chi2 test is a widely used hypothesis testing method. As one of the nonparametric test methods, it can be used to detect the correlation between independent variables and dependent variables. By testing the correlation between the features and the classification labels of the training samples, the importance of features can be sorted, and then feature selection can be realized.

ReliefF

ReliefF, an extension of the Relief method, decomposes the multi-classification problem into several two classification problems (Robnik-Sikonja and Kononenko 2003). In order to realize feature selection, the ReliefF algorithm comprehensively considers within-class distance and between-class distance of samples to update the weights of features. If the minimum within-class distance is greater than the minimum between-class distance, the feature weight should be increased; otherwise, the weight should be reduced. Then, the features are sorted according to the weight. The feature with smaller weight has less contribution to accurate classification. The weight of feature t is shown in Equation 2:

$$W_t^{class(x_i)} = W_t^{class(x_{i-1})} + \frac{1}{n}\left\{\frac{\sum_{c\neq class}^{class(x)} D_t(x_i, M(x_i))}{(n-1)m_{class(x)}} - \frac{D_t(x_i, H(x_i))}{m_{class(x)}}\right\} \tag{2}$$

where x_i represent the i-th sample in class x; n is the number of classes; $m_{class(x)}$ is the number of samples in class x; $D_t(x_i,M(x_i))$ is the distance between x_i and the nearest sample in same class; $D_t(x_i,H(x_i))$ is the distance between x_i and the nearest sample in different class.

Recursive Feature Elimination

Recursive feature elimination, as a greedy algorithm, has at its core idea of the multi-round training model (Liu and An 2020). Firstly, all features are used to train the model, and each feature gets an initial weight. After that, the features with smaller weight are removed, and the model is retrained by using the remaining features, and each feature gets a new weight. After multi-round training until all the features are traversed, the order in which the features are removed can reflect the priority of the features.

Random Forest

Random forest is composed of a large number of classification regression decision trees, which is an fused machine learning algorithm with good robustness and simple use (Breiman 2001). Random forest randomly extracts samples and features and uses multiple decision trees for training to prevent over fitting effectively. Each node of the decision tree is a condition about a certain feature in order to divide the data set into two parts according to different response variables. GINI is a measurement for node impurity which indicates how often a random instance will be misclassified. The rationale behind it is that important features can significantly reduce the impurity in samples. In the random forest algorithm, the average reduced GINI impurity of each feature can be used as the criterion for feature selection.

Extra Tree

Extra tree is also a fused machine learning algorithm. Similar to the random forest, it is also composed of a large number of decision trees. Different from random forest, each decision tree in extra tree uses all samples for training; only features are selected randomly. Extra tree are a set of "free growing" decision trees (Geurts *et al.* 2006) in which the nodes of decision trees are bifurcated based on completely random values. Compared with random forest, extra tree has stronger randomness and higher generalization ability. In the extra tree algorithm, the sorting and selection of features are also realized based on GINI impurity.

Table 2. Description of the data sources and extracted features.

Feature category	Layer	Source	Description
Band features	1-12	Sentinel-2	B1 (Coastal aerosol); B2 (Blue); B3 (Green); B4 (Red); B5/B6/B7 (Vegetation Red Edge); B8 (NIR); B8A (Vegetation Red Edge); B9 (Water Vapour); B11/B12 (SWIR)
	13-14	Sentinel-1	B13 (VV BSC); B14 (VH BSC)
Index features	15-26	Sentinel-2	NDVI: (B8-B4)/(B8+B4); NDWI: (B3-B8)/(B3+B8); NDBI: (B11-B8)/(B11+B8); RVI: B8/B4; SAVI: (B8-B4)(1+0.5)/(B8+B4+0.5); RNDVI: (B5-B4)/(B5+B4); REDNDVI: (B8-B5)/(B8+B5); TVI: 0.5((120(B8-B3))-(200(B4-B3))); RRI1: B8/B5; RRI2: B5/B4; MSRre: (B8/B5-1)/($\sqrt{\text{B8/B5+1}}$); CIre: B7/B5-1
	27-30	Sentinel-1	Ratio: B13/B14; Difference: B13-B14 Sum: B13+B14; RDS: (B13-B14)/(B13+B14)
Textural features	31-38/ 39-46	Sentinel-2/ Sentinel-1	Mean; Variance; Homogeneity; Contrast; Dissimilarity; Entropy; Second Moment; Correlation

Hyperparameter Optimization SVM

Based on the principle of structural risk minimization, SVM improves the generalization ability of the classification model, thus minimizing the experience risk and confidence interval. In addition, SVM can also obtain good statistical rules from a small number of training samples. The core idea of the SVM classification model is to find an optimal hyperplane to maximize the separation of samples in different classes.

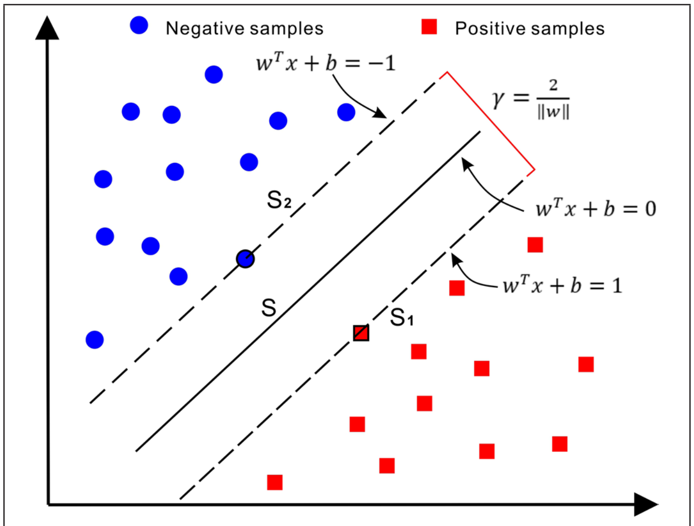


Figure 3. The optimal hyperplane in support vector machine model.

Suppose there is a training samples set $D=\{(x_1,y_1),(x_2,y_2),\ldots,(x_N,y_N)\}$, where $x_i \in R^m$, $y_i \in \{-1,1\}$ ($i=1,2,\ldots,N$). Red and green dots represent positive samples ($y = 1$) and negative samples ($y = -1$), respectively. These two classes of samples are linearly separable (Figure 3). S is the hyperplane between two classes of samples represented by $f(x)=w^Tx+b$, where w is the dimension coefficient vector and b is the offset. S_1 and S_2 represent the planes nearest to the hyperplane passing through the positive sample and negative sample, respectively. The distance between S_1 and S_2 is the classification interval. SVM determines the optimal hyperplane by solving the following optimization problem (Equation 3):

$$max \frac{2}{\|w\|}, s.t. \, y_i\left(w^T x_i + b\right) \geq 1, i = 1,2,\ldots,N \tag{3}$$

According to the Lagrange multiplier method, the problem of finding the optimal hyperplane can be transformed into a convex quadratic programming problem (Equation 4):

$$L(w,b,\alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{N} \alpha_i \left(y_i\left(w \bullet x_i + b\right) - 1\right) \tag{4}$$

where α_i is a Lagrange multiplier and $\alpha_i \geq 0$. The final classification discriminant function is shown in Equation 5:

$$f(x) = sign\left(\sum_{i=1}^{N} \alpha_i y_i \left(x \bullet x_i\right) + b\right) \tag{5}$$

In many instances, the training samples set is not linearly separable. SVM uses kernel function to map training samples into high-dimensional space, and then designs optimal hyperplane to make training samples linearly separable (Figure 4). The kernel function $K(\,)$ uses a nonlinear transformation to replace the inner product between two samples. In this case, the classification discriminant function is shown in Equation 6:

$$f(x) = sign\left(\sum_{i=1}^{N} \alpha_i y_i K\left(x, x_i\right) + b\right) \tag{6}$$

In this study, the widely used linear kernel function and radial basis function were selected for SVM models training. The parameters need to be set including C and γ, where C is the penalty coefficient, indicating the tolerance of misclassification and γ is used to determine the distribution of training samples mapped to high-dimensional space.

In order to obtain the optimal land cover mapping results, grid search and five-fold cross-validation were used to realize the hyperparameter optimization of SVM model; that is, to obtain the optimal combination of parameters C and γ. Grid Search is a process of selecting different C and γ combinations at a certain interval within a predefined range. For different combinations of C and γ, cross-validation is used to test the accuracy of classification results by rotation estimation of training samples.

Accuracy Assessment

Combined with high spatial resolution Google Earth images, 600 pixels were selected as validation samples by means of random stratified sampling, i.e., 126 pixels form impervious surface, 88 pixels from bare land, 88 pixels from grass, 108 pixels from forest, 101 pixels from water, and 89 pixels from shadow area (Figure 5). Kappa coefficient (KA), overall

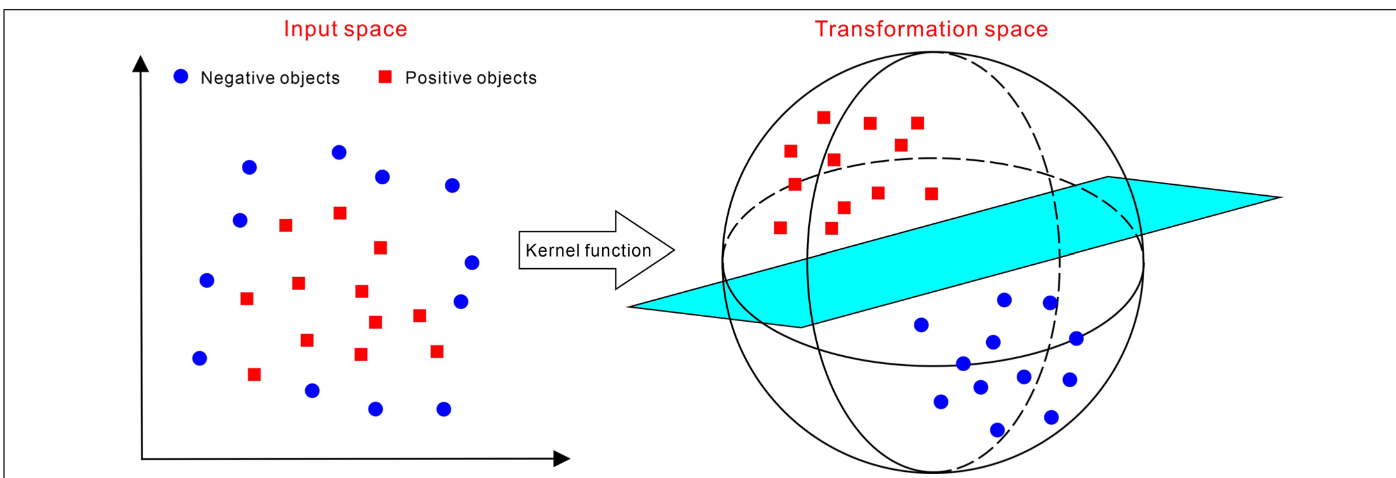


Figure 4. Training samples mapped to high dimensional space by kernel function.

accuracy (OA), producer's accuracy (PA) and user's accuracy (UA) calculated based on confusion matrix were used to assess the accuracy of the mapping results.

The confusion matrix can be obtained by comparing the ground-truth class with the predicted class of verification samples. Each row in the confusion matrix represents a ground truth class, and the values in the row correspond to the prediction results of the ground truth pixels. KA is one of the most important assessment parameters, which is calculated as shown in Equation 7. As a consistency test index, it is used to measure quality of mapping results. When the KA is less than 0.75, the classification model has relatively poor performance.

$$KA = \frac{N\sum_{i=1}^{n} X_{ii} - \sum_{i=1}^{n}(X_{i+} \times X_{+i})}{N^2 - \sum_{i=1}^{n}(X_{i+} \times X_{+i})} \tag{7}$$

where n represent the number of classes; N represent the number of validation samples; X_{ii} stands for the element of the i-th row and the i-th column in the confusion matrix; X_{i+} and X_{+i} represent the sum of the elements in i-th row and the i-th column, respectively. OA is the ratio of the number of correctly classified pixels to the total number of invalidation pixels, which is used to evaluate the overall performance of classification model.

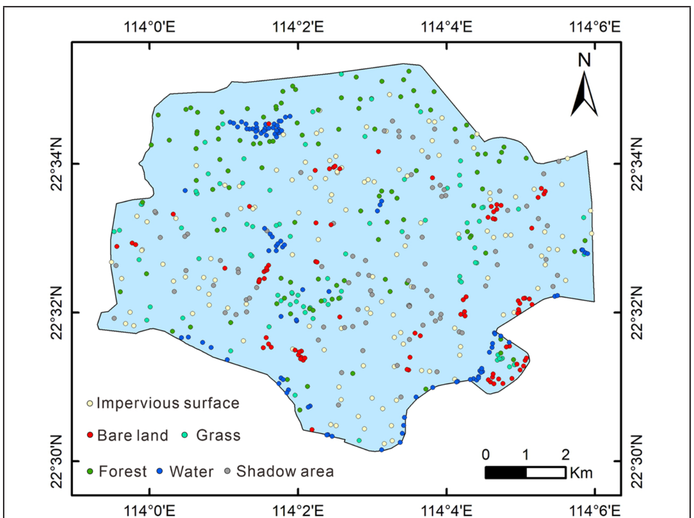


Figure 5. Distribution of validation samples in the study area.

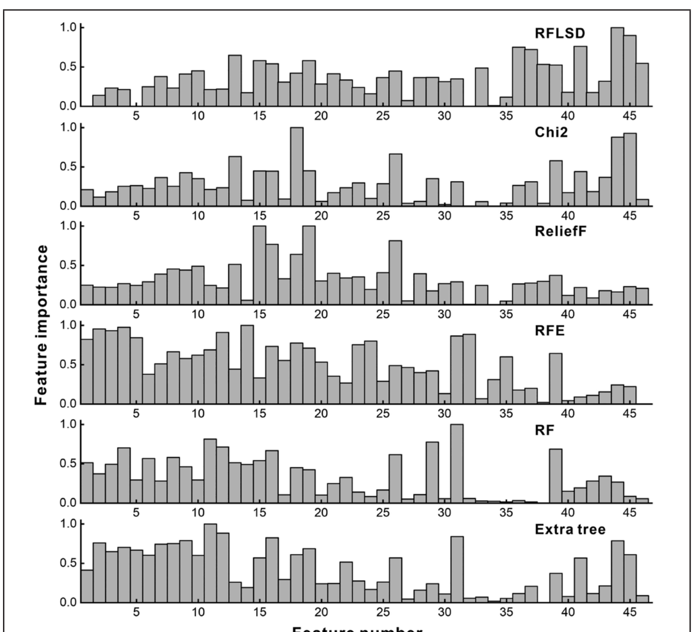


Figure 6. Summary of normalized feature importance scores obtained by six feature selection methods.

Results

Feature Importance Ranking and Subset Selection

In this study, 46 features were extracted from Sentinel-1 and Sentinel-2 data and applied to land cover mapping in Futian District, including the original band features, index features and texture features. Using the six feature selection methods introduced in the section "Feature Selection", the importance of 46 features was evaluated from the perspectives of information theory, mathematical statistics, classification, etc. (Figure 6). Feature importance scores obtained by different methods have been normalized between 0 and 1. The distribution of feature importance scores obtained by six methods is different. In contrast, the feature importance ranking results obtained by RFE, RF and Extra tree methods have high consistency.

The purpose of feature selection is to remove features that are irrelevant to or have negative impact on accurate mapping to improve the processing efficiency and maintain or improve the performance of the model. According to the importance scores ranking of features obtained by different methods, 24 features (nearly half of the original number) were selected from the 46 extracted features to form the feature subsets. Six feature subsets were input into hyperparameter optimization SVM model as training data for land cover mapping so as to reduce the data dimension and verify the effect of feature selection. The feature subsets and the ranking of feature importance scores obtained by different methods are shown in Table 3.

Table 3. Feature importance score ranking, and subset composition obtained by different methods.

Rank	RFLSD	Chi2	ReliefF	RFE	RF	Extra tree
1	F44	**F18**	**F19**	F14	**F31**	F11
2	F45	F45	F15	F4	F11	F12
3	F41	F44	**F26**	F2	F29	**F31**
4	F36	**F26**	**F16**	F3	F12	**F16**
5	F37	F13	**F18**	F12	F4	**F9**
6	F13	**F39**	F13	F32	**F39**	F44
7	**F19**	**F19**	**F10**	**F31**	**F16**	F2
8	F15	F15	F8	F5	**F26**	F8
9	F46	**F16**	**F9**	F1	F8	**F7**
10	**F16**	F41	F25	F24	F6	F4
11	F38	**F9**	F21	**F18**	F15	**F19**
12	**F39**	F43	F28	F23	F13	F5
13	F33	**F7**	**F7**	**F16**	F1	F3
14	**F10**	**F10**	**F39**	**F19**	F3	F45
15	**F26**	F29	F23	F11	F14	**F18**
16	**F18**	F37	F22	F8	**F9**	F6
17	F21	**F31**	F17	**F39**	**F18**	**F10**
18	**F9**	F23	F20	**F10**	**F19**	F15
19	**F7**	F25	F38	F35	F2	**F26**
20	F29	F36	F6	**F9**	F43	F41
21	F25	F5	**F31**	F17	F22	F22
22	F28	F4	F37	F20	**F10**	F1
23	**F31**	F8	F4	**F7**	F5	**F39**
24	F22	F12	F30	**F26**	**F7**	F17

Note: F44: feature 44 (Sentinel-1 Entropy). The features in bold refer to the features existing in all six feature subsets.

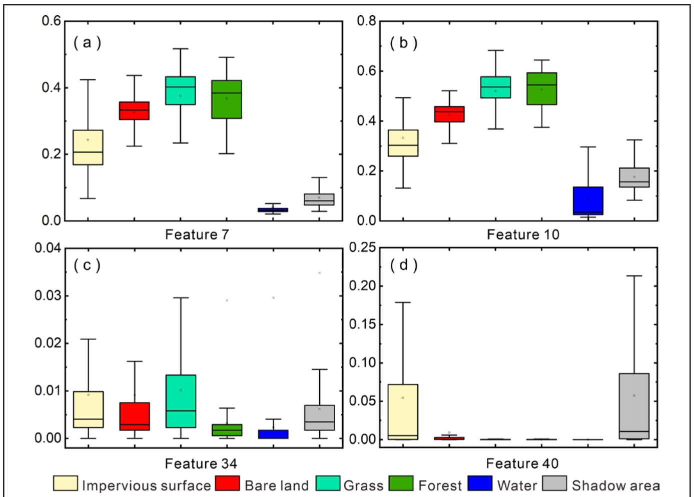


Figure 7. Data distribution of each land cover in features 7, 10, 34, and 40.

It can be seen from Table 3 that the composition of feature subsets and the ranking of feature importance scores selected by the six methods were quite different. According to the statistical results, several features were the common features of the six feature subsets, such as feature 7 (Vegetation Red Edge), feature 10 (Water Vapor), feature 19 (soil-adjusted vegetation index (SAVI)), and feature 39 (Sentinel-1 Mean). While feature 27 (Sentinel-1 Ratio), feature 34 (Sentinel-2 Contrast), feature 40 (Sentinel-1 Variance), and feature 42 (Sentinel-1 Contrast) did not appear in any feature subset. Therefore, this study further selected features 7, 10, 34, and 40 as representatives to explore the data distribution of each land cover and preliminarily verify the reliability of feature selection (Figure 7).

From Figure 7, it can be found that the data span range of feature 7 and feature 10 was large, and the data distribution overlap degree of different land covers was small. Land cover types can be well distinguished based on these features. However, the data span of feature 34 and feature 40 was relatively small, and the data distribution of different land covers was highly overlapped, so it was not easy to distinguish land covers based on feature 34 and feature 40. Therefore, feature selection can effectively eliminate features that contribute less to accurate mapping or are unfavorable to accurate mapping. That means, select the features with high discrimination for different land covers to obtain better mapping results.

Comparison of ULC Mapping Results Corresponding to Different Feature Subsets.

In order to evaluate the role of feature selection and compare the effect difference of different methods, six feature subsets (24 features) and all-feature set (46 features) were respectively taken as training data and input into the hyperparameter optimization SVM for ULC mapping in Futian District. The confusion matrixes, overall accuracies and kappa coefficients are shown in Table 4.

Among the classification models corresponding to six feature selection methods, the recursive feature elimination-support vector machine (RFE-SVM) model had the highest OA and KA, which were 89.17% and 0.8695, respectively. It suggested that RFE-SVM model performed better than others. The classification accuracy of each model from high to low was RFE-SVM, RF-SVM, Extra tree-SVM, RFLSD-SVM, ReliefF-SVM, and Chi2-SVM. Notably, the OA and KA of all-feature SVM model were only 81.33% and 0.7742, which were lower than those of any feature subset SVM model. This again showed that feature selection can effectively eliminate features irrelevant to mapping and reduce data noise, thus improving the accuracy of mapping results. Compared with all-feature SVM model, the mapping results of RFE-SVM model (Figure 8) were greatly improved in terms of the OA and KA, which were increased by 7.84% and 0.0953, respectively.

According to the confusion matrixes (Table 4), the misclassification of pixels in the land cover mapping results mainly occurred between the classes with high spectral mixing, such as impervious surface, bare land and grass. In addition to the mapping results of RFE-SVM model, a large number of bare land pixels were misclassified as impervious surface. In contrast, RFE-SVM model had a strong ability to extract bare land, which greatly decreased the incidence of

Table 4. The confusion matrixes, overall accuracies and kappa coefficients of mapping results.

Land cover classes predicted by SVM	IS	BL	G	F	W	SA
RFLSD-SVM						
IS	115	4	1	0	0	6
BL	30	57	0	0	1	0
G	13	0	66	5	0	4
F	6	0	10	90	0	2
W	1	0	2	0	94	4
SA	14	0	0	0	1	74
OA:82.67%	KA:0.7900					
ReliefF-SVM						
IS	113	5	1	0	0	7
BL	26	61	0	0	0	1
G	19	0	53	12	0	4
F	8	0	4	96	0	0
W	2	0	3	0	89	7
SA	11	0	0	0	0	78
OA:81.67%	KA:0.7778					
RF-SVM						
IS	107	8	3	0	0	8
BL	10	76	0	0	1	1
G	11	0	62	11	0	4
F	6	0	1	97	0	4
W	1	0	2	1	93	4
SA	10	0	0	1	1	77
OA:85.33%	KA:0.8229					
All-feature SVM						
IS	107	14	0	0	0	5
BL	17	70	0	0	1	0
G	16	0	52	16	0	4
F	10	0	8	90	0	0
W	1	0	4	0	93	3
SA	12	0	0	0	1	76
OA:81.33%	KA:0.7742					
Chi2-SVM						
IS	108	11	1	0	0	6
BL	20	67	0	0	0	1
G	14	0	61	10	0	3
F	9	0	9	89	0	1
W	2	0	5	0	88	6
SA	12	0	0	0	1	76
OA:81.50%	KA:0.7763					
RFE-SVM						
IS	106	12	3	0	0	5
BL	7	78	1	0	1	1
G	4	0	72	8	0	4
F	6	0	2	98	0	2
W	2	0	1	0	98	0
SA	5	0	0	0	1	83
OA:89.17%	KA:0.8695					
Extra tree-SVM						
IS	108	7	6	0	0	5
BL	17	69	1	0	1	0
G	6	0	69	9	0	4
F	7	0	4	93	0	4
W	1	0	3	0	93	4
SA	10	0	0	1	1	77
OA:84.83%	KA:0.8169					

Note: IS: Impervious surface; BL: Bare land; G: Grass; F: Forest; W: Water; SA: Shadow area.

this misclassification phenomenon. However, the ability of the RFE-SVM model to extract impervious surface was below average. 12, 3, and 5 impervious surface pixels were wrongly classified as bare land, grass, and shadow area, respectively. To further analyze the extraction ability of different models for each land cover, the PA and UA of different land covers was obtained.

Combining Table 4 and Figure 9, it can be seen that the omission error of bare land and grass was relatively large. Forest and water had a smaller classification error and better extraction effect. The highest PA was 91.27% for impervious surface (RFLSD), 88.64% for bare land (RFE), 81.82% for grass (RFE), 90.74% for forest (RFE), 97.03% for water (RFE), and 93.26% for shadow area (RFE). The highest UA was 81.54% for impervious surface (RFE), 93.44% for bare land (RFLSD), 91.18% for grass (RF), 94.74% for forest (RFLSD), 100.00% for water (ReliefF), and 87.37% for shadow area (RFE).

In addition, the average PA and UA of different land covers in the RFE-SVM mapping results were the highest, which were 89.27% and 89.53%, respectively. Except for the impervious surface, the PA of other classes had reached the leading level. This showed that RFE-SVM model can balance the mapping effect of different land covers and was more suitable for land cover mapping in urban area with high spectral and spatial complexity. The average PA and UA of the mapping results obtained by the all-feature SVM model were 80.73% and 83.23%, which were also lower than the results obtained by feature subset SVM models. This proved again the validity of using feature selection strategy to improve the accuracy of classification model.

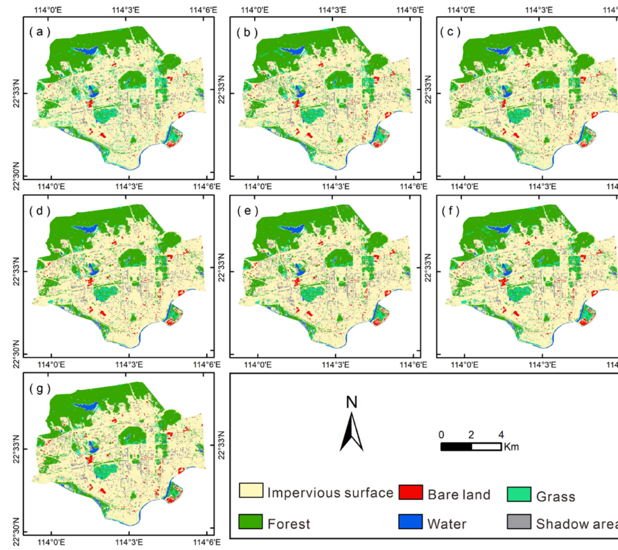


Figure 8. Urban land cover mapping results obtained from (a) removing features with low standard deviation (RFLSD)-support vector machine (SVM); (b) Chi2-SVM; (c) ReliefF-SVM; (d) recursive feature elimination-support vector machine (RFE-SVM); (e) random forest (RF)-SVM; (f) Extra tree-SVM; (g) all-feature SVM.

Discussion

Analysis of Feature Correlations

Correlation analysis is a statistical analysis method used to measure the degree of closeness between two features. The increase of the absolute value of the average correlation in remote sensing data may lead to the occurrence of data redundancy, and thus reduce the interpretability of the model. To verify the validity of this view, the Pearson correlation coefficient between each pair of 46 features was calculated (Figure 10). Obviously, the first five spectral bands (feature 1 to feature 5) of Sentinel-2 data were highly correlated with each other. And there were also high correlation coefficients between features 6, 7, 8, 9, 10, and feature 31. Moreover, note that there were also some highly correlated pairs in the index features (feature 15 to feature 26) derived from the Sentinel-2 data. However, the correlation between texture features and other features was low, and there were some highly independent feature pairs.

In order to clarify the influence of correlation on model accuracy, the distribution of the absolute value of correlation between features in different feature subsets was further obtained (Figure 11). On the whole, the distribution of feature correlation in the six feature subsets was similar to that in the all-feature set. In other words, the absolute value of correlation of most feature pairs was between 0 and 0.6. It should be noted that in the interval [0.8, 1], the frequency of absolute correlation coefficient in the all-feature set was only 9.38%, while that in each feature subset was more than 15%, with the highest frequency being 21.82% (RF and Extra tree). This indicated that, compared with the all-feature set, the proportion of highly correlated feature pairs in the feature subsets obtained by feature selection was larger.

Statistical results showed that the average absolute value of correlation coefficient of feature pairs in the all-feature set was 0.3531. The average absolute value of correlation coefficient in each feature subset were 0.5047 (Extra tree), 0.4910

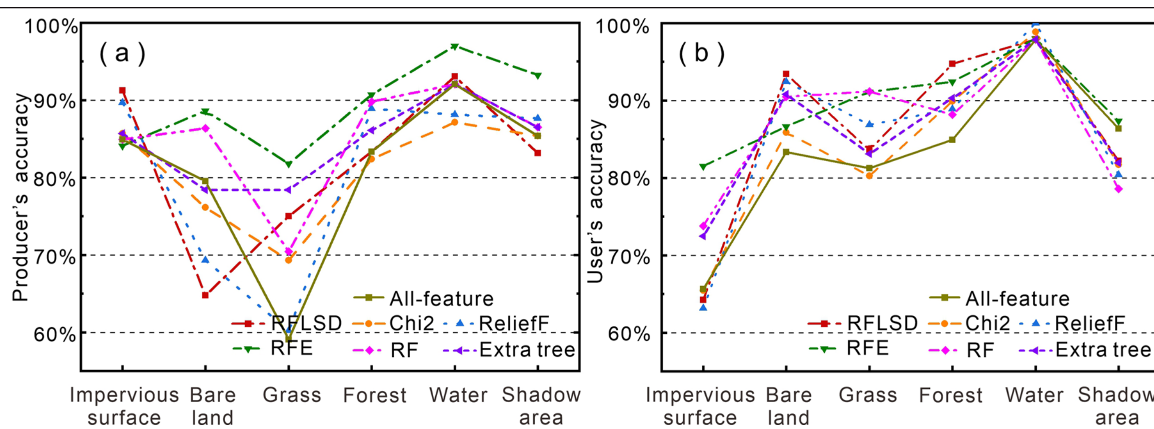


Figure 9. The producer's accuracy (a) and the user's accuracy (b) of each land cover from mapping results based on different feature sets.

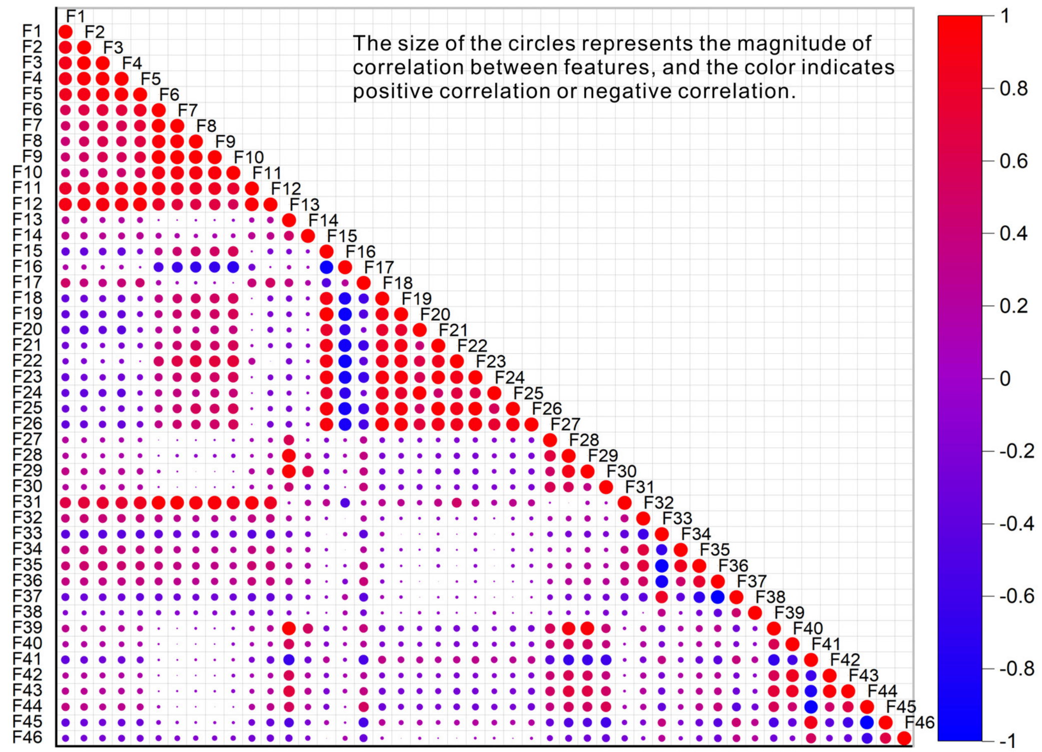


Figure 10. Results of the Pearson correlation coefficients of the 46 extracted features. Note: F1: feature 1. Check Table 2 for more details on the 46 extracted features.

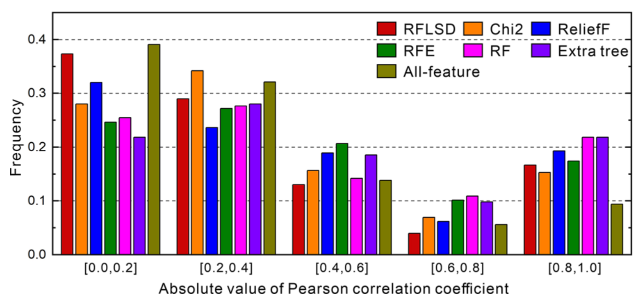


Figure 11. Frequency distribution of absolute value of the feature correlation coefficient.

(RF), 0.4801 (RFE), 0.4586 (ReliefF), 0.4324 (Chi2), and 0.4036 (RFLSD) from high to low. After feature selection, the average absolute value of correlation coefficient of feature pairs in each feature subset was higher than that in the all-feature set.

Feature selection reduced the dimension of the data set, and at the same time caused a small increase in average absolute value of correlation coefficient of feature pairs. This increase was due to the removal of irrelevant features and the retention of features that contribute more to accurate mapping. In addition, the section "Comparison of ULC Mapping Results Corresponding to Different Feature Subsets" showed that feature selection can effectively improve the accuracy of land cover mapping results based on hyperparameter optimization SVM model. This showed that the correlation increase caused by feature selection not only did not affect the interpretability of the model, but also played a positive role in improving the accuracy of land cover mapping results. Combined with the mapping results in the section "Comparison of ULC Mapping Results Corresponding to Different Feature Subsets", it was found that there was a positive correlation (see Equation 8) between the average absolute value of correlation coefficient and the overall accuracy of land cover mapping results, and the correlation coefficient R was 0.6408.

$$y = 0.3415x + 0.6855 \tag{8}$$

where x represents the average absolute value of correlation coefficient in each feature set, and y represents the overall accuracy of mapping results based on SVM model when each feature set was used as training data.

Validation of Multi-source Remote Sensing Data Fusion Applied to ULC Mapping

In this study, the Sentinel-2 original bands (feature 1 to feature 12) and the Sentinel-1 backscattering coefficients (features 13, 14) were used as training data, and the land cover mapping results (Figure 12) in Futian District were obtained again based on the hyperparameter optimization SVM model. The corresponding confusion matrixes, overall accuracies and kappa coefficients are shown in Table 5. By comparing the differences between the mapping results, the effect of multi-feature fusion of optical and SAR data in improving the land cover mapping was verified.

The OA and KA of land cover mapping results obtained by using Sentinel-2 data were 81.00% and 0.7700, respectively. The confusion matrix showed that Sentinel-2 data had good extraction ability for forest and water. Sentinel-2 data had the worst ability to extract grass, and the PA was only 62.50%. There were 12, 19 and 2 grass pixels misclassified as impervious surface, forest, and shadow area. The land cover mapping results obtained only by using the Sentinel-1 backscattering coefficients were poor, and the OA and KA were only 48.53% and 0.3364. Among all kinds of land cover, Sentinel-1 data had better recognition effect on water and impervious surface.

According to the section "Comparison of ULC Mapping Results Corresponding to Different Feature Subsets", through the fusion of optical and SAR data and feature selection, the OA and KA of land cover mapping results in Futian District were up to 89.17% and 0.8695 (RFE-SVM model). Compared with the mapping results obtained by using Sentinel-2 data

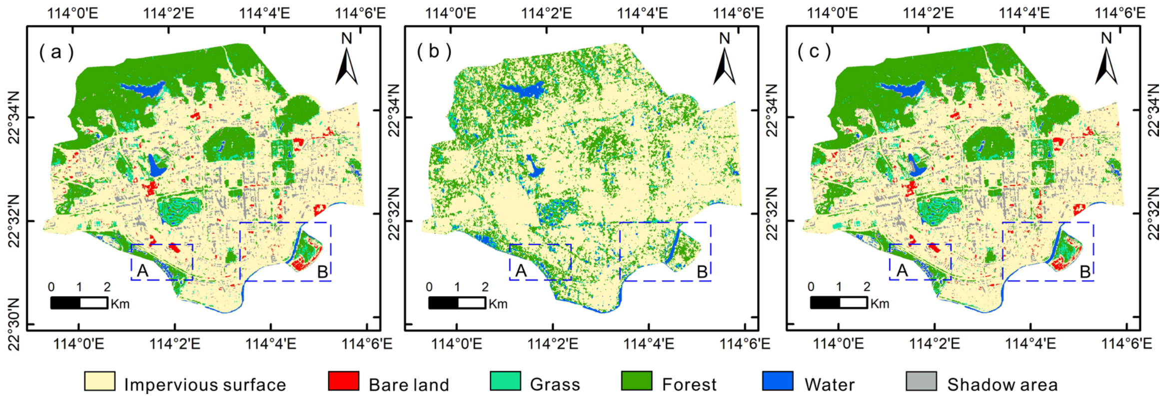


Figure 12. Urban land cover mapping results based on hyperparameter optimization support vector machine model. (a) Sentinel-2 original bands (feature 1 to feature 12); (b) Sentinel-1 backscattering coefficients (features 13, 14); (c) Sentinel-2 and Sentinel-1 fusing data (recursive feature elimination method).

alone, the OA and KA of the mapping results obtained from multi-source data were improved by 8.17% and 0.0995, respectively. Through the fusion of optical and SAR remote sensing data, the land mapping results with higher accuracy were obtained. The misclassification between water and shadow area, grass and forest caused by spectral mixing was greatly reduced.

In addition, two different regions (A and B) in Futian District were selected for visual comparison of land cover mapping results. The locations and land cover mapping results of two regions are shown in Figure 12 and Figure 13, respectively.

Region A is located in the southwestern corner of Futian District. Compared with the three land cover mapping results, it can be found that the accurate identification of roadside trees (marked with a red circle) cannot be achieved only by using Sentinel-2 band features. Some forest pixels were mistakenly classified as impervious surface. The accuracy of land cover mapping results obtained by using Sentinel-1 data was low, and the degree of patch fragmentation was high. Using multi-source remote sensing data, the mapping results based on RFE-SVM model achieved better recognition of roadside trees, and the overall mapping effect had been significantly improved.

Region B is located in the southeast corner of Futian District and contains a river (marked with a red circle). Due to spectral confusion, some water pixels were mistakenly classified as grass in Sentinel-2 mapping results. Sentinel-1 data can effectively identify the water, but it did not accurately identify any bare land pixels. In contrast, the fusion of multi-source remote sensing data realized the accurate identification of water and bare land, and effectively reduced the occurrence of misclassification and omission.

Table 5. Confusion matrixes, overall accuracies and kappa coefficients of mapping results obtained from Sentinel-2 and Sentinel-1 data.

Land cover classes predicted by SVM						
	IS	BL	G	F	W	SA
Sentinel-2 (feature 1 to feature 12)						
IS	108	8	6	0	0	4
BL	16	70	0	0	1	1
G	12	0	55	19	0	2
F	8	0	2	98	0	0
W	6	0	3	0	85	7
SA	15	0	1	1	2	70
PA(%)	85.71	79.55	62.50	90.74	84.16	78.65
UA(%)	65.45	89.74	82.09	83.05	96.59	83.33
OA:81.00% KA:0.7700						
Sentinel-1 (features 13, 14)						
IS	104	0	7	12	3	-
BL	70	0	5	9	4	-
G	32	0	6	46	4	-
F	51	0	2	55	0	-
W	3	0	10	5	83	-
SA	-	-	-	-	-	-
PA(%)	82.54	0	6.82	50.93	82.18	-
UA(%)	40.00	-	20.00	43.31	88.30	-
OA:48.53% KA:0.3364						

Conclusions

High precision land covers data can play an important role in urban planning, management, and development. However, due to the spectral complexity and spatial heterogeneity, it is challenging to obtain accurate ULC mapping results using optical remote sensing data only. In order to improve ULC mapping, this study proposed a mapping framework that fused optical and SAR data and optimized feature subsets. Sentinel-2 and Sentinel-1 data were fused, and multiple feature selection methods were evaluated to reduce data dimension and noise to realize high-precision land cover mapping in Futian District based on hyperparameter optimization SVM model. The main conclusions are as follows:

(1) Feature selection can eliminate features with low contribution or negative influence to accurate mapping, reduce model complexity, and improve land cover mapping results. Based on the hyperparameter optimization SVM model, the OA and KA of the mapping results obtained from feature subsets selected by different methods were higher than those obtained from the all-feature set. Among them, the RFE-SVM model had the highest accuracy, with the OA and KA reaching 89.17% and 0.8695, respectively. Compared with the mapping results obtained from the all-feature set, the OA and KA were improved by 7.84% and 0.0953, respectively.

(2) Feature selection can reduce data dimension, but it can also lead to a slight increase in the average absolute value of correlation coefficient between features. This increase was caused by the elimination of features unrelated to the accurate mapping. It not only did not influence the interpretability of the model, but also played a positive role in improving

the accuracy of land cover mapping. The statistical results showed that there was a positive correlation between the average absolute value of correlation coefficient and the overall accuracy of land cover mapping results, with correlation coefficient R reaching 0.6408.

(3) Multi-feature fusion of multi-source remote sensing data can significantly improve the effectiveness of land cover mapping. Compared with the land cover mapping results obtained by using Sentinel-2 or Sentinel-1 data alone, the results obtained based on RFE-SVM model were improved by 8.17% and 40.64% in OA, and 0.0995 and 0.5331 in KA. It is found from the specific comparison regions that the fusion of optical and SAR data can effectively reduce the confusion between impervious surface and forest, water and grass, and effectively improve the land cover mapping in complex urban areas.

Acknowledgements
This work was supported in part by the National Key R&D Program of China under Grant 2018YFB2100501; in part by the National Natural Science Foundation of China under Grants 42090012, 41771452, 41771454; in part by the Key R&D Program of Yunnan Province in China under Grant 2018IB023.

Conflicts of Interest
The authors declare no conflict of interest.

References

Breiman, L. 2001. Random forests. *Machine Learning* 45(1):5–32.

Cai, J., J. Luo, S. Wang and S. Yang. 2018. Feature selection in machine learning: A new perspective. *Neurocomputing* 300:70–79.

Chen, J., L. Chen, F. Chen, Y. Ban, S. Li, G. Han, X. Tong, C. Liu, V. Stamenova and S. Stamenov. 2021a. Collaborative validation of GlobeLand30: Methodology and practices. *Geo-spatial Information Science* 24(1):134–144.

Chen, L., F. Rottensteiner and C. Heipke. 2021b. Feature detection and description for image matching: From hand-crafted design to deep learning. *Geo-Spatial Information Science* 24(1):58–74.

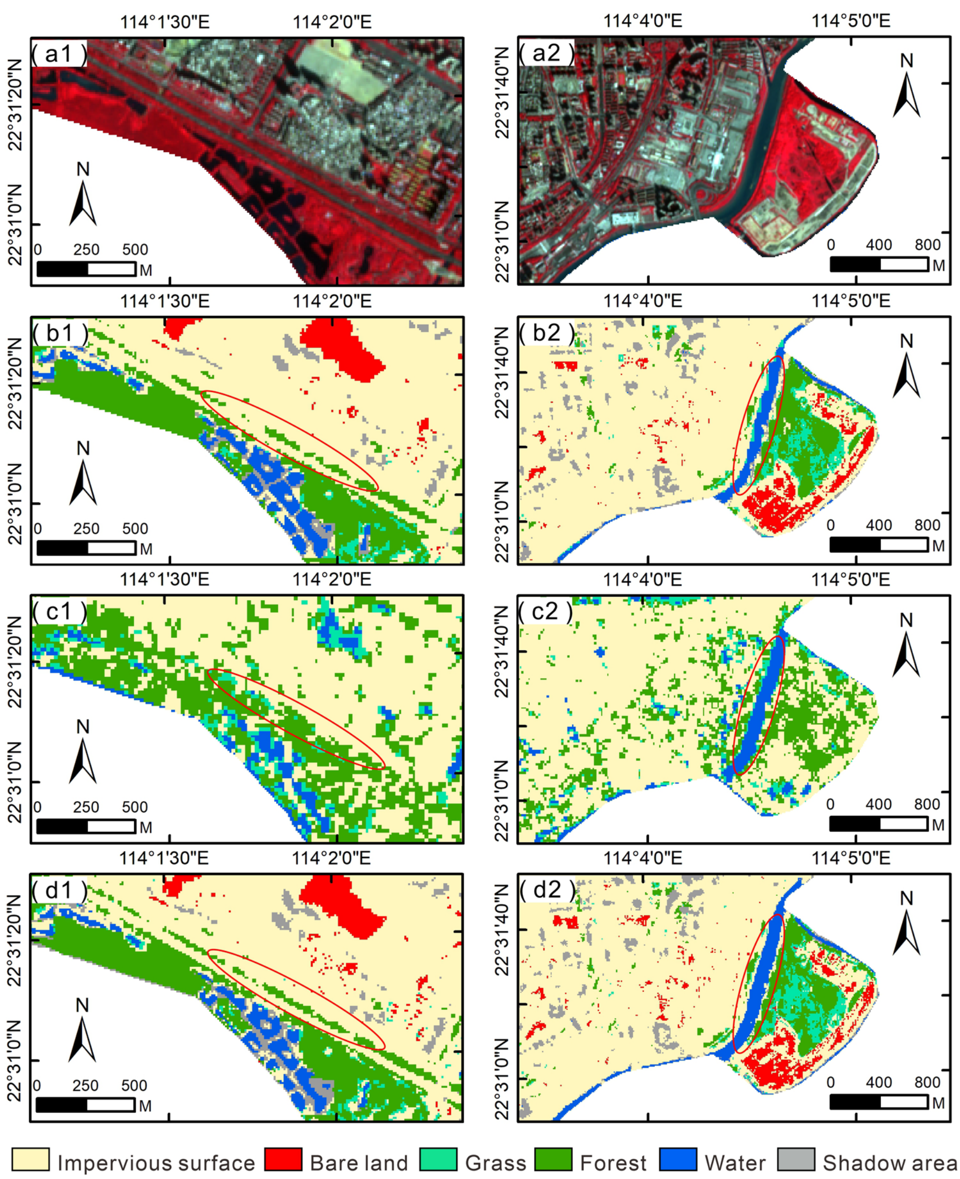


Figure 13. False color composites of a Sentinel-2 image covering the two comparison regions (a1: region A; a2: region B) and the corresponding land cover mapping results from (b) Sentinel-2 data, (c) Sentinel-1 data, (d) Sentinel-2 and Sentinel-1 fusing data (recursive feature elimination-support vector machine (RFE-SVM) model).

Clinton, N., L. Yu and P. Gong. 2015. Geographic stacking: Decision fusion to increase global land cover map accuracy. *ISPRS Journal of Photogrammetry and Remote Sensing* 103:57–65.

Cortes, C. 1995. Support-vector networks. *Machine Learning* 20(3):273–297.

Foody, G. M. and A. Mathur. 2006. The use of small training sets containing mixed pixels for accurate hard image classification: Training on mixed spectral responses for classification by a SVM. *Remote Sensing of Environment* 103(2):179–189.

Friedl, M., D. McIver, J. Hodges, X. Zhang, D. Muchoney, A. H. Strahler, C. Woodcock, S. Gopal, A. Schneider, A. Cooper, A. Baccini, F. Gao and C. Schaaf. 2002. Global land cover mapping from MODIS: Algorithms and early results. *Remote Sensing of Environment* 83(1):287–302.

Gallego, F. J. 2004. Remote sensing and land cover area estimation. *International Journal of Remote Sensing* 25(15):3019–3047.

Georganos, S., T. Grippa, S. Vanhuysse, M. Lennert, M. Shimoni, S. Kalogirou and E. Wolff. 2018. Less is more: Optimizing classification performance through feature selection in a veryhigh- resolution remote sensing object-based urban application. *GIScience and Remote Sensing* 55(2):221–242.

Geurts, P., D. Ernst and L. Wehenkel. 2006. Extremely randomized trees. *Machine Learning*, 63(1):3–42.

Griffiths, P., C. Nendel and P. Hostert. 2019. Intra-annual reflectance composites from *Sentinel-2* and Landsat for national-scale crop and land cover mapping. *Remote Sensing of Environment* 220:135–151.

Guyon, I. M. and A. Elisseeff. 2003. An introduction to variable and feature selection. *The Journal of Machine Learning Research* 1157–1182.

Hodgson, M. E., J. R. Jensen, J. A. Tullis, K. D. Riordan and C. M. Archer. 2003. Synergistic use of Lidar and color aerial photography for mapping urban parcel imperviousness. *Photogrammetric Engineering and Remote Sensing* 69(9):973–980.

Huang, C., L. S. Davis and J. Townshend. 2002. An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing* 23(4):725–749.

Huang, B. and J. Wang. 2020. Big spatial data for urban and environmental sustainability. *Geo-spatial Information Science* 23(2):125–140.

Joshi, N., M. Baumann, A. Ehammer, R. Fensholt, K. Grogan, P. Hostert, M. R. Jepsen, T. Kuemmerle, P. Meyfroidt, E.T.A. Mitchard, J. Reiche, C. M. Ryan and B. Waske. 2016. A review of the application of optical and radar remote sensing data fusion to land use mapping and monitoring. *Remote Sensing* 8(1):70.

Khatami, R., G. Mountrakis and S. V. Stehman. 2016. A metaanalysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research. *Remote Sensing of Environment* 177:89–100.

Kuang, W., Y. Dou, C. Zhang, W. Chi, A. Liu, Y. Liu, R. Zhang and J. Liu. 2015. Quantifying the heat flux regulation of metropolitan land use/land cover components by coupling remote sensing modeling with in situ measurement. *Journal of Geophysical Research-Atmospheres* 120(1):113–130.

Lazzarini, M., A. Molini, P. Marpu, T. Ouarda and H. Ghedira. 2015. Urban climate modifications in hot desert cities: The role of land cover, local climate, and seasonality. *Geophysical Research Letters* 42(22), 9980–9989.

Li, J., C. Song, L. Cao, F. Zhu, X. Meng and J. Wu. 2011. Impacts of landscape structure on surface urban heat islands: A case study of Shanghai, China. *Remote Sensing of Environment* 115(12):3249–3263.

Li, Y., Z. Cao, H. Long, Y.-s. Liu and W. Li. 2017. Dynamic analysis of ecological environment combined with land cover and NDVI changes and implications for sustainable urban–rural development: The case of Mu Us Sandy Land, China. *Journal of Cleaner Production* 142:697–715.

Liu, H. and H. An. 2020. Preliminary tests on the performance of MLC-RFE and SVM-RFE in Lansat-8 image classification. *Arabian Journal of Geosciences* 13(3).

Pal, M. 2005. Random forest classifier for remote sensing classification. *International Journal of Remote Sensing* 26(1):217–222.

Prins, A. J. and A. V. Niekerk. 2020. Crop type mapping using LiDAR, Sentinel-2 and aerial imagery with machine learning algorithms. *Geo-Spatial Information Science* 24(2):215–227.

Robnik-Sikonja, M. and I. Kononenko. 2003. Theoretical and empirical analysis of ReliefF and RReliefF. *Machine Learning* 53:23–69.

Sesnie, S. E., P. E. Gessler, B. Finegan and S. Thessler. 2008. Integrating Landsat TM and SRTM-DEM derived variables with decision trees for habitat classification and change detection in complex neotropical environments. *Remote Sensing of Environment* 112(5):2145–2159.

Shao, Z., H. Fu, P. Fu and L. Yin. 2016. Mapping urban impervious surface by fusing optical and SAR data at the decision level. *Remote Sensing* 8(11):945.

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* 12:5620–5620.

Shao, Z., W. Wu and S. Guo. 2020b. IHS-GTF: A fusion method for optical and synthetic aperture radar data. *Remote Sensing* 12(17):2796.

Shih, H.-C., D. a. Stow and Y. h. Tsai. 2019. Guidance on and comparison of machine learning classifiers for Landsat-based land cover and land use mapping. *International Journal of Remote Sensing* 40(4):1248–1274.

Simonetti, E., D. Simonetti and D. Preatoni. 2014. *Phenology-Based Land Cover Classification Using Landsat 8 Time Series*. Luxembourg: Publications Office of the European Union.

Sukawattanavijit, C., J. Chen and H. Zhang. 2017. GA-SVM algorithm for improving land-cover classification using SAR and optical remote sensing data. *IEEE Geoscience and Remote Sensing Letters* 14(3):284–288.

Symeonakis, E., T. P. Higginbottom, K. Petroulaki and A. Rabe. 2018. Optimisation of savannah land cover characterisation with optical and SAR data. *Remote Sensing* 10(4):499.

Tabib Mahmoudi, F., A. Arabsaeedi and S. K. Alavipanah. 2019. Feature-level fusion of Landsat 8 data and SAR texture images for urban land cover classification. *Journal of the Indian Society of Remote Sensing* 47(3):479.

Ty, H., J. Yang, L. Xuecao and P. Gong. 2016. Mapping urban land use by using Landsat images and open social data. *Remote Sensing* 8(2):151.

Weng, Q., X. Hu and H. Liu. 2009. Estimating impervious surfaces using linear spectral mixture analysis with multitemporal ASTER images. *International Journal of Remote Sensing* 30(18):4807–4830.

Zhou, Y., R. Zhang, S. Wang and F. Wang. 2018. Feature selection method based on high-resolution remote sensing images and the effect of sensitive features on classification accuracy. *Sensors* 18(7):2013.

Zhang, F. and X. Yang. 2020. Improving land cover classification in an urbanized coastal area by random forests: The role of variable selection. *Remote Sensing of Environment* 251:112105.

Zhang, H., J. Li, T. Wang, H. Lin, Z. Zheng, Y. Li and Y. Lu. 2018. A manifold learning approach to urban land cover classification with optical and radar data. *Landscape and Urban Planning* 172:11–24.

Zhang, H. and R. Xu. 2018. Exploring the optimal integration levels between SAR and optical data for better urban land cover mapping in the Pearl River Delta. *International Journal of Applied Earth Observations and Geoinformation* 64:87–95.

Zhang, Y. and L. Sun. 2019. Spatial-temporal impacts of urban land use land cover on land surface temperature: Case studies of two Canadian urban areas. *International Journal of Applied Earth Observations and Geoinformation* 75:171–181.

Zhu, Z., C. Woodcock, J. Rogan and J. Kellndorfer. 2012. Assessment of spectral, polarimetric, temporal, and spatial dimensions for urban and peri-urban land cover classification using Landsat and SAR data. *Remote Sensing of Environment* 117:72–82.

Examining the Integration of Landsat Operational Land Imager with Sentinel-1 and Vegetation Indices in Mapping Southern Yellow Pines (Loblolly, Shortleaf, and Virginia Pines)

Clement E. Akumu and Eze O. Amadi

Abstract

The mapping of southern yellow pines (loblolly, shortleaf, and Virginia pines) is important to supporting forest inventory and the management of forest resources. The overall aim of this study was to examine the integration of Landsat Operational Land Imager (OLI) optical data with Sentinel-1 *microwave C-band satellite data and vegetation indices in mapping the canopy cover of southern yellow pines. Specifically, this study assessed the overall mapping accuracies of the canopy cover classification of southern yellow pines derived using four data-integration scenarios: Landsat OLI alone; Landsat OLI and* Sentinel-1*; Landsat OLI with vegetation indices derived from satellite data—normalized difference vegetation index, soil-adjusted vegetation index, modified soil-adjusted vegetation index, transformed soil-adjusted vegetation index, and infrared percentage vegetation index; and 4) Landsat OLI with* Sentinel-1 *and vegetation indices. The results showed that the integration of Landsat OLI reflectance bands with* Sentinel-1 *backscattering coefficients and vegetation indices yielded the best overall classification accuracy, about 77%, and standalone Landsat OLI the weakest accuracy, approximately 67%. The findings in this study demonstrate that the addition of backscattering coefficients from* Sentinel-1 *and vegetation indices positively contributed to the mapping of southern yellow pines.*

Introduction

Southern yellow pines such as loblolly pine (*Pinus taeda*), Virginia pine (*P. virginiana*), and shortleaf pine (*P. echinata*) are softwood forest vegetation species commonly found in the southeastern United States. These pine species are commercially marketed and provide economic benefits to the country. For example, loblolly and shortleaf pines are usually grown for pulpwood and sawlogs, whereas Virginia pine is usually grown as Christmas-tree species (English *et al.* 2004; Young *et al.* 2007).

The mapping of softwood forest vegetation species such as loblolly, shortleaf, and Virginia pines is important for effective management of forest resources (Xie *et al.* 2008; Ke *et al.* 2010; Deng *et al.* 2011; Shang and Chisholm 2014; Roth *et al.* 2015). For example, updated digital maps of forest vegetation species and canopy cover are continually being sought by forest managers and policy makers to support management decisions and policies (Skidmore *et al.* 1997; Rozenstein and Karnieli 2011). Furthermore, forest vegetation canopy cover maps can help to understand tree-species ecology for community dynamics as well as species inputs into the ecosystems (van Ewijk *et al.* 2014). They can also be used as inputs for modeling and other forest management and planning activities such as harvesting, regeneration, and fire management (van Aardt and Wynne 2007; Hamilton *et al.* 2021).

The spectral information of satellite remotely sensed data, such as Landsat Operational Land Imager (OLI) optical data and *Sentinel-1* C-band synthetic-aperture radar (SAR) sensor data, make them feasible and cost-effective in mapping forest vegetation canopy cover compared to traditional field-survey methods over large geographic areas (Xie *et al.* 2008; Shang and Chisholm 2014; Vincent *et al.* 2019). However, because many individually sensed images have either high spatial resolution or high spectral resolution, there is a need to integrate satellite remotely sensed data to improve image classification. For example, Jiménez *et al.* (2017) and Fatoyinbo and Armstrong (2010) integrated Landsat Enhanced Thematic Mapper Plus with lidar and National Forest Inventory data to map aboveground forest cover and biomass, and found a more accurate estimation of aboveground forest biomass using this data-integration method. Wan *et al.* (2021) integrated multispectral *Sentinel-2* image data with high-spatial-resolution aerial images for tree-species classification of forest stands. They classified and mapped 11 forest vegetation species stands and found an increase in overall mapping accuracy after data integration. Furthermore, Biswas *et al.* (2020) evaluated the contribution of three satellite data sources—Landsat OLI, *Sentinel-1*, and *Sentinel-2*—in mapping diverse forest vegetation types in Myanmar. They found that using a combination of *Sentinel-1* and *Sentinel-2* data produced the highest accuracy (89.6%), followed by *Sentinel-2* alone (87.97%) and Landsat OLI (82.68%).

Satellite-derived vegetation indices are useful indicators of forest biophysical condition and can be integrated with satellite remotely sensed data to further improve the discrimination of forest vegetation and canopy cover. This is because spectral vegetation indices measure the photosynthetic size of plant canopies. Furthermore, they are used as indicators to monitor variations in temporal and spatial characteristics of vegetation structure and density (Xue and Su 2017; Akumu *et al.* 2021). For example, Prabhakara *et al.* (2015)

Clement E. Akumu and Eze O. Amadi are with the Department of Agricultural and Environmental Sciences, College of Agriculture,

Tennessee State University, Nashville, TN 37209-1561 (aclemen1@tnstate.edu).

Contributed by Sidike Paheding, May 3, 2021 (sent for review July 15, 2021; reviewed by Jorge A.S. Centeno, Cary Cox).

Photogrammetric Engineering & Remote Sensing
Vol. 88, No. 1, January 2022, pp. 29–38.
0099-1112/22/29–38

doi: 10.14358/PERS.21-00024R2

used vegetation indices to ascertain the correlation between vegetation biomass, ground cover, and derived indices in Maryland (USA). They found a strong correlation between the normalized difference vegetation index (NDVI) and percent vegetation cover. Furthermore, they found the triangular vegetation index most accurate in estimating vegetation biomass. Bera *et al.* (2020) used vegetation indices such as the NDVI, advanced vegetation index, shadow index, and bareness index to detect and monitor forest vegetation canopy cover and health. They found a reduction in forest canopy cover and density between 1998 and 2009 in the Silabati River Basin (India). Furthermore, Reid *et al.* (2016) generated the NDVI from Landsat Thematic Mapper data as an indicator of forest productivity to examine forest cover and health trends at Fort Benning, Georgia. They found that most plots had declining greenness through time, consistent with the overall NDVI trend.

Other recent studies have integrated vegetation indices derived from satellite data with remotely sensed satellite data to map forest canopy cover and habitats (Martinuzzi *et al.* 2008; Sinha *et al.* 2015; Abdollahnejad *et al.* 2019; Ganz *et al.* 2020). For example, Sinha *et al.* (2015) integrated the thermal integrated vegetation index and advanced thermal integrated vegetation index with Landsat Enhanced Thematic Mapper Plus satellite data to map land cover including forest canopy cover in a semi-arid deciduous forest landscape. They found that the classification accuracy of land cover improved with integration of the thermal vegetation indices from the Landsat Enhanced Thematic Mapper Plus thermal band with spectral information. Rhyma *et al.* (2020) integrated Satellite pour l'observation de la Terre (*SPOT-6* and *SPOT-7*) satellite data with the NDVI and soil-adjusted vegetation index (SAVI) to discriminate forest canopy cover. They found satellite data-derived vegetation indices useful in improving the accuracy of classification in a mangrove forest ecosystem. Although satellite-derived vegetation indices and satellite data have been integrated in forest canopy cover classification, there is no known knowledge of the integration of Landsat OLI optical data with *Sentinel-1* C-band SAR sensor data and derived vegetation indices for mapping forest canopy cover of southern yellow pines. The integration of Landsat OLI optical data with *Sentinel-1* microwave satellite data and derived vegetation indices could improve the overall detection, mapping, and classification accuracy of the canopy cover of southern yellow pines.

The overall aim of this study is to examine the integration of Landsat OLI optical data with *Sentinel-1* C-band SAR sensor satellite data and derived vegetation indices in mapping the canopy cover of southern yellow pines (loblolly, shortleaf, and Virginia pines). Specifically, this study assesses the overall mapping accuracies of the canopy cover classification of southern yellow pines derived using four data-integration scenarios: Landsat OLI alone; Landsat OLI and *Sentinel-1*; Landsat OLI with satellite data-derived vegetation indices—NDVI, SAVI, modified soil-adjusted vegetation index (MSAVI), transformed soil-adjusted vegetation index (TSAVI), infrared percentage vegetation index (IPVI); and Landsat OLI with *Sentinel-1* and satellite data-derived vegetation indices. To the best of our knowledge, this is the first study to examine the integration of satellite data-derived vegetation indices with Landsat OLI optical and *Sentinel-1* C-band SAR sensor data in the classification and mapping of the canopy cover of southern yellow pines.

Materials and Methods

Study Area

Marion County, Tennessee, was selected as a case study area in this study (Figure 1). It is located between latitude 35.319 492 34°N and 34.984 474 18°N, and between longitude 85.361 694 34°W and 85.872 871 40°W. The county is in the southern region of Tennessee and occupies approximately 516 mi^2 of surface area. This study area was selected because of the availability of cloud-free Landsat OLI satellite data and several field data sets of southern yellow pines.

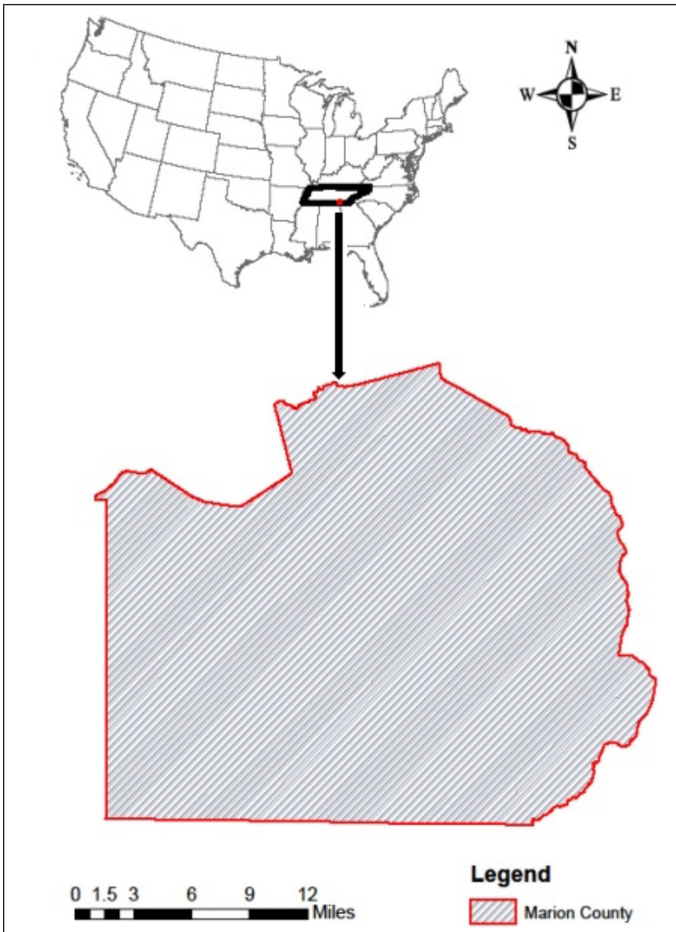


Figure 1. Study area: Marion County, Tennessee, United States of America.

Vegetation

A significant part of the study area is covered by forest vegetation, especially softwood forest vegetation such as southern yellow pines. Southern yellow pines commonly found in the region included loblolly pine (*P. taeda*), Virginia pine (*P. virginiana*), and shortleaf pine (*P. echinata*). In addition to softwood forest vegetation, there is also hardwood forest vegetation in the area, with common species including locust (*Gleditsia* spp.), poplar (*Populus* spp.), maple (*Acer* spp.), oak (*Quercus* spp.), elm (*Ulmus* spp.), hickory (*Carya* spp.), and sycamore (*Platanus* spp.; Akumu *et al.* 2018).

Climate

The climate of the region is characterized by hot summers and moderately cold winters with some erratic cold spells and snowfall (Akumu *et al.* 2018; Hodges *et al.* 2018). The seasonal average temperatures are 41°F in the winter, 60°F in the spring, 78°F in the summer, and 60°F in the fall (Hinkle 1989). The mean annual temperature of Marion County is about 78°F. Average precipitation in the region is about 51 in. (1300 mm), evenly distributed over the seasons (Hodges *et al.* 2018).

Geology

Marion County is on the Cumberland Plateau and contains a good portion of Sequatchie Valley and part of the Tennessee River. The plateau is formed by level rocks. The tableland of the Cumberland Plateau, Walden Ridge, and the Raccoon Mountain crest are capped by sandstones, shales, conglomerates, and coal seams (Hodges *et al.* 2018). The Tennessee and Sequatchie River floors are made of limestones of Ordovician and Mississippian origin which contain alkaline soils (Akumu *et al.* 2018). The most noticeable landform in the county is Sequatchie Valley, which runs northeast to southwest through the center of the county. The valley is linear and covers about 25% of the total area of the county (Starnes 1986).

Methodology

The methodology for this study involved six data-processing: acquisition of Landsat OLI optical data and *Sentinel-1* microwave satellite data; preprocessing of satellite data; generation of satellite-data vegetation indices; data integration; classification of the canopy cover of southern yellow pines; and validation/accuracy assessment (Figure 2).

The Landsat OLI satellite data, with an acquisition date of 28 February 2016, were downloaded from the United States Geological Survey website (http://earthexplorer.usgs.gov) as a Level-1 cloud-free scene. Landsat OLI satellite data have 11 spectral bands, with a spatial resolution of 30 m for bands 1–7 and 9 (Table 1). Bands 1–7 were used in the classification and mapping of the canopy cover of southern yellow pines.

This study selected a Landsat OLI satellite data set with a winter acquisition date because southern yellow pines are conifers that are easily detected in the winter season, when deciduous trees shed their leaves. The Landsat OLI scene with 30-m spatial resolution was subsetted for the study area and geometric correction was performed. The geometric correction was carried out using more than 50 ground control points with a root-mean-square (RMS) error < 1 pixel. The RMS error is the distance between the input (source) location of a ground control point and the transformed location of the same ground control point (Tawfeik *et al.* 2016). Using more than 50 ground control points is acceptable if the RMS error is < 1 pixel, but unacceptable if it is > 1 pixel (Nguyen 2015; Pehani *et al.* 2016; Tawfeik *et al.* 2016). This is because an RMS error < 1 pixel provides a high-quality georeferenced image compared to an RMS error > 1 pixel (Baboo *et al.* 2011; Tawfeik *et al.* 2016).

Radiometric correction was performed on the Landsat OLI satellite data by converting digital numbers to at-surface reflectance. It entails correcting image pixel values for variation in the sun elevation angle and calibrating images to account for degradation of the sensor over time. Changes in sensor calibration factors will obscure real changes on the ground

Table 1. Landsat Operational Land Imager spectral bands and characteristics.

Band	Wavelength (μm)	Resolution (m)
1: Ultra Blue (coastal/aerosol)	0.43–0.45	30
2: Blue	0.45–0.51	30
3: Green	0.53–0.59	30
4: Red	0.64–0.67	30
5: Near-infrared	0.85–0.88	30
6: Shortwave infrared 1	1.57–1.65	30
7: Shortwave infrared 2	2.11–2.29	30
8: Panchromatic	0.50–0.68	15
9: Cirrus	1.36–1.38	30
10: Thermal infrared 1	10.60–11.19	100 × 30
11: Thermal infrared 2	11.50–12.51	100 × 30

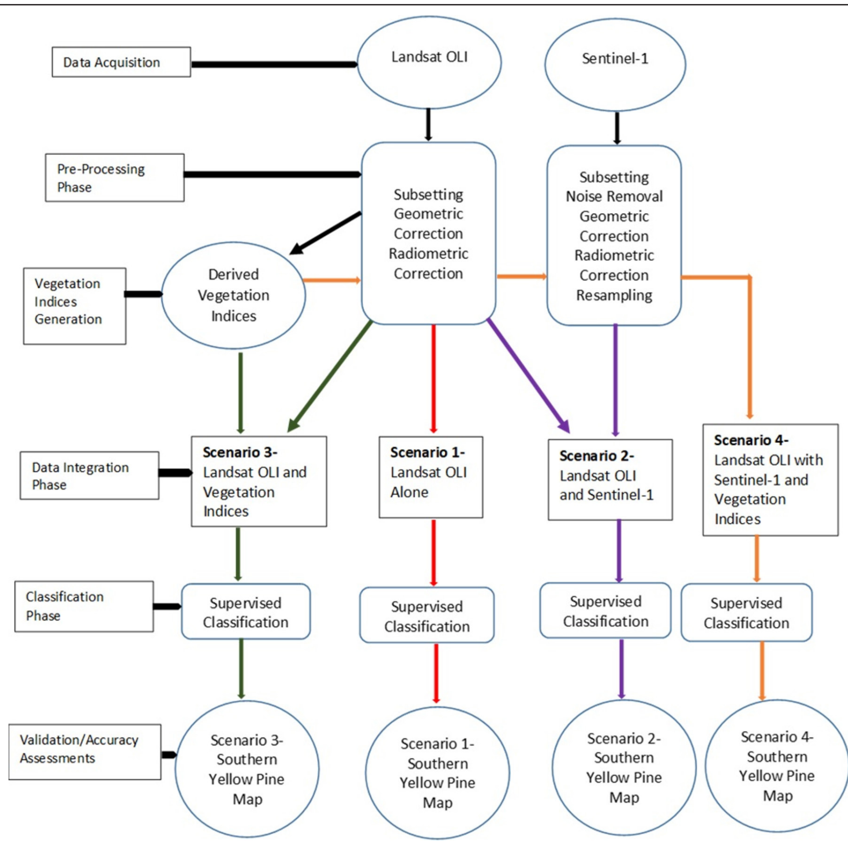


Figure 2. A schematic representation of the methodology used to classify southern yellow pines using four data-integration methods.

(Mather and Koch 2011). The Landsat OLI scene was converted from digital numbers to at-surface reflectance by using reflectance rescaling coefficients derived by the United States Geological Survey (2019):

$$\rho\lambda' = M_p Q_{\text{cal}} + A_p \tag{1}$$

where $\rho\lambda'$ = top-of-atmosphere (TOA) planetary reflectance without correction for solar angle, M_p = band-specific multiplicative rescaling factor (Reflectance_Mult_Band_*x*, where *x* is the band number), A_p = band-specific additive rescaling factor (Reflectance_Add_Band_*x*), and Q_{cal} = digital numbers. The factors Reflectance_Mult_Band_*x* and Reflectance_Add_Band_*x* were obtained from the header file of the imagery.

Furthermore, the correction of TOA planetary reflectance for sun angle was performed using the equation (United States Geological Survey 2019)

$$\rho\lambda = \rho\lambda'/\sin(\theta_{\text{SE}}) \tag{2}$$

where $\rho\lambda$ =TOA planetary reflectance corrected for sun angle, $\rho\lambda'$ = TOA planetary reflectance without correction for solar angle, and θ_{SE} = local sun elevation angle (in degrees), provided in the metadata (Sun_Elevation).

The *Sentinel-1* C-band SAR sensor satellite data, with an acquisition date of 24 January 2018, were downloaded from the European Space Agency Data Hub (https://scihub.copernicus.eu/dhus/#/home) as a Sentinel-1A scene. *Sentinel-1* has a C-band with four acquisition modes: Stripmap, Interferometric Wide swath, Extra Wide swath, and Wave (Table 2). The Interferometric Wide swath mode vertical-vertical, vertical-horizontal, horizontal-vertical, and horizontal-horizontal polarizations was used in the classification and mapping of the canopy cover of southern yellow pines.

Table 2. Mode, spectral resolution, swath, and polarization of Sentinel-1 C-band SAR sensor.

Mode	Incidence Angle (°)	Resolution (m)	Swath Width (km)	Polarization
Stripmap	20–45	5×5	80	HH+HV, VH+VV, HH, VV
Interferometric Wide swath	29–46	5×20	250	HH+HV, VH+VV, HH, VV
Extra Wide swath	19–47	20×40	400	HH+HV, VH+VV, HH, VV
Wave	22–35 35–38	5×5	20×20	HH, VV

H = horizontal; V = vertical.

The *Sentinel-1* microwave scene with a spatial resolution of 5×20 m was subsetted to the study area and noise removal (speckle filtering) was performed. The noise removal was carried out using spatial averaging in a 60×60-m window. Geometric correction was performed on the scene using the Shuttle Radar Topography Mission global digital elevation map for the study area. The digital elevation map was used to provide terrain correction, and the *Sentinel-1* data were reprojected to the WGS84 - UTM Zone 16 map projection. Radiometric correction was performed on the imagery by converting the digital numbers to backscattering coefficients (σ^o; Twele *et al.* 2016):

$$\sigma^o = \frac{\text{DN}^2 \sin\theta}{K} \tag{3}$$

where θ = incidence angle, K = calibration constant, and DN = digital numbers.

The backscattering coefficients were then expressed in decibels (Twele *et al.* 2016):

$$\sigma^o{}_{dB} = 10log_{10}(\sigma^o) \tag{4}$$

The scene was later resampled to the same spatial resolution as the Landsat OLI satellite data—30-m cell size.

The TOA reflectance image of the Landsat OLI satellite data was used to generate the NDVI, SAVI, MSAVI, TSAVI, and IPVI. These indices were selected because they are indicators of plant greenness and are considered to take into account the effect of soil background. For example, these indices have a spectral red band that is strongly absorbed by plant chlorophyll and is an indicator of vegetation greenness. Furthermore, they also have an infrared band that is strongly absorbed when plants become stressed by factors such as dehydration, lack of nutrients, diseases, and leaf-structure deterioration (Qi *et al.* 1994; Lichtenthaler *et al.* 1996). In addition, the TSAVI has an adjustment factor to minimize the effect of soil background (Baret *et al.* 1989). The vegetation indices were also selected because they can be easily generated from the Landsat OLI spectral bands and could contribute to the discrimination of southern yellow pines. Other indices, such as the normalized difference water index and modified normalized difference water index, were not considered because they have shortwave infrared bands and are good indicators of vegetation wetness rather than greenness (Gao 1996; Xu 2006).

The normalized difference vegetation index was generated as (Lichtenthaler *et al.* 1996)

$$\text{NDVI} = \frac{\text{Near-infrared} - \text{Red}}{\text{Near-infrared} + \text{Red}} \tag{5}$$

The soil-adjusted vegetation index was generated using (Huete 1988)

$$\text{SAVI} = \frac{(\text{Near-infrared} - \text{Red})(1 + L)}{(\text{Near-infrared} + \text{Red} + L)} \tag{6}$$

where L is the soil brightness conversion factor of 0.5.

The modified soil-adjusted vegetation index was generated using (Qi *et al.* 1994)

$$\text{MSAVI} = \frac{(\text{Near-infrared} - \text{Red})(1 + L)}{\text{Near-infrared} + \text{Red} + L} \tag{7}$$

where L is calculated by

$$L = 1 - \frac{2 * s * (\text{Near-infrared} - \text{Red}) * (\text{Near-infrared} - s * \text{Red})}{(\text{Near-infrared} + \text{Red})} \tag{8}$$

in which s is the slope of the soil line from a plot of brightness values of red versus near-infrared.

The transformed soil-adjusted vegetation index was generated using (Baret *et al.* 1989)

$$\text{TSAVI} = \frac{a(\text{NIR} - a * \text{Red} - b)}{\left[\text{Red} + a(\text{NIR} - b) + 0.08\left(1 + a^2\right)\right]} \tag{9}$$

where a and b are the slope and intercept of the soil line, respectively; 0.08 is the adjusted coefficient value; and NIR is the near-infrared value.

The infrared percentage vegetation index was generated using (Crippen 1990):

$$\text{IPVI} = \frac{\text{Near-infrared}}{\text{Near-infrared} + \text{Red}} \tag{10}$$

First, the stand-alone Landsat OLI reflectance scene was used to classify and map the canopy cover of southern yellow pines, as scenario 1. The Landsat OLI visible and infrared spectral bands were used in classifying the canopy cover of southern yellow pines. Second, the Landsat OLI reflectance scene was integrated with *Sentinel-1* backscattering coefficients to classify and map southern yellow pines, as scenario 2. Third, the Landsat OLI reflectance scene was integrated with the derived vegetation indices NDVI, SAVI, MSAVI, TSAVI, and IPVI to classify and map southern yellow pines, as scenario 3. Fourth, the Landsat OLI reflectance scene was integrated with *Sentinel-1* backscattering coefficients and derived vegetation indices to classify and map southern yellow pines, as scenario 4. The spectral bands of the Landsat OLI reflectance scene were integrated directly as separate bands with *Sentinel-1* backscattering coefficients and derived vegetation indices.

Supervised classification was performed to classify and map the canopy cover of southern yellow pines in all four data-integration scenarios. The canopy cover of loblolly, shortleaf, and Virginia pines was classified and mapped using training data from 22 field sites. The southern yellow pines on each field site covered a large geographic area of >200,000 m^2. The sites represented homogenous stands of loblolly, shortleaf, and Virginia pines. Most of the southern yellow pines at the field sites were at least 6 m tall. There were seven sites of loblolly pine, 12 of shortleaf, and three 3 of Virginia pine. The site-location data were obtained from area foresters at the Tennessee Department of Agriculture. Sixty polygons (20 loblolly, 20 shortleaf, and 20 Virginia pine) were digitized from the 22 field sites to serve as training data in the supervised classification process.

The supervised classification was performed using a machine-learning random-forest classification algorithm, with the 60 digitized polygons of southern yellow pines serving as training data. The random-forest classification model was controlled for overfitting by five-fold cross-validation repeated twice on the training data. During the cross-validation process, about 25% of the training data were kept aside as test data set. The remaining 75%—the training data set—was divided into five equal sets and used in the five-fold cross-validation. The first set was kept as the holdout (testing) set and the remaining sets were used to train the random-forest classification prediction model of southern yellow pines. The five-fold cross-validation was performed with a changing holdout (testing) set. The mean accuracy of the canopy cover classification of southern yellow pines generated from the five-fold cross-validation process was estimated. The training data were then used in the random-forest classification of the canopy cover of southern yellow pines, and the kept-aside 25% test data set was used to validate the classification. The accuracy with the test and training data sets was then evaluated (Sharma *et al.* 2017; Costa *et al.* 2018; Elmaz *et al.* 2020).

Furthermore, the numbers of trees and training samples in the random-forest classification prediction model were selected through a resampling-based procedure to search for optimal tuning parameters. The optimal settings were selected based on the mean overall accuracy across the five-fold cross-validation, repeated twice (Sharma *et al.* 2017; Costa *et al.* 2018). The default number of training samples was selected and set at 5000, and the number of random-forest trees was set at 10. The random-forest classification algorithm was selected because it has been found to outperform other machine-learning classification algorithms such as support vector machines in mapping forest canopy cover and species (Shang and Chisholm 2014; Sharma *et al.* 2017; Elmahdy *et al.* 2020; Sjöqvist *et al.* 2020).

The canopy cover maps of southern yellow pines generated using the four data-integration classification methods were validated to examine how well they represented southern yellow pines on the ground. The validation effort was performed by randomly selecting 100 polygons from each classified canopy cover map. The validation data (100 polygons) were distinct from the training data (60 polygons) used in the random-forest classification of the canopy cover of southern yellow pines.

Determination of ground truth by field-plot visits and use of Google Earth Pro information was used to validate the classified canopy cover maps derived from the four data-integration scenarios. The overall accuracy was computed for each classified map by dividing the total correct (the sum of the major diagonal in the error matrix table) by the total number of pixels in the error matrix table (Mather and Koch 2011). The κ coefficient was also measured as described by Mather and Koch (2011). The classified canopy cover maps were later exported into Geographic Information System for extent analyses.

Results and Discussion

The canopy cover of southern yellow pines representing loblolly, shortleaf, and Virginia pines (Figures 3–6) was successfully classified and mapped using the four data-integration classification methods. The distribution of loblolly, shortleaf, and Virginia pines was similar in all four scenarios. The canopy cover of shortleaf pine was more intense in the northern parts of the study area than the southern parts. Similarly, the canopy cover of loblolly and Virginia pines was more abundant in the northern parts of the study area than the southern portions. The lesser canopy cover of southern yellow pines in the southern parts of the study area is likely because of intense harvesting. Southern yellow pines are continually harvested as pulpwood and saw timber products in the region (Clabo and Clatterbuck 2005; Hansen *et al.* 2014). Furthermore, on average, shortleaf pine had the most canopy cover with all four data-integration classification methods, and Virginia pine had the least canopy cover (Table 3). The dry, better-drained ridgetops associated with the Cumberland Plateau, which are commonly found in the region, possibly provided suitable conditions for growing shortleaf pines (Hodges *et al.* 2018).

The overall, user, and producer accuracies varied in all data-integration scenarios. The overall accuracy is the average of the individual class accuracies expressed as a percentage (Mather and Koch 2011). The user accuracy is a measure of how well the classified canopy cover of loblolly, shortleaf, and Virginia pines on the map represented southern yellow pines on the ground. The producer accuracy is the ability of the random-forest classification algorithm to detect southern yellow pines.

Table 3. Percentage canopy cover of loblolly, shortleaf, and Virginia pines derived with the four data-integration classification methods.

Southern Yellow Pine	Scenario 1: Landsat OLI Alone	Scenario 2: Landsat OLI and *Sentinel-1* Data	Scenario 3: Landsat OLI and Vegetation Indices	Scenario 4: Landsat OLI with *Sentinel-1* Data and Vegetation Indices
Loblolly	14	17	23	14
Shortleaf	71	73	62	73
Virginia	15	10	15	13

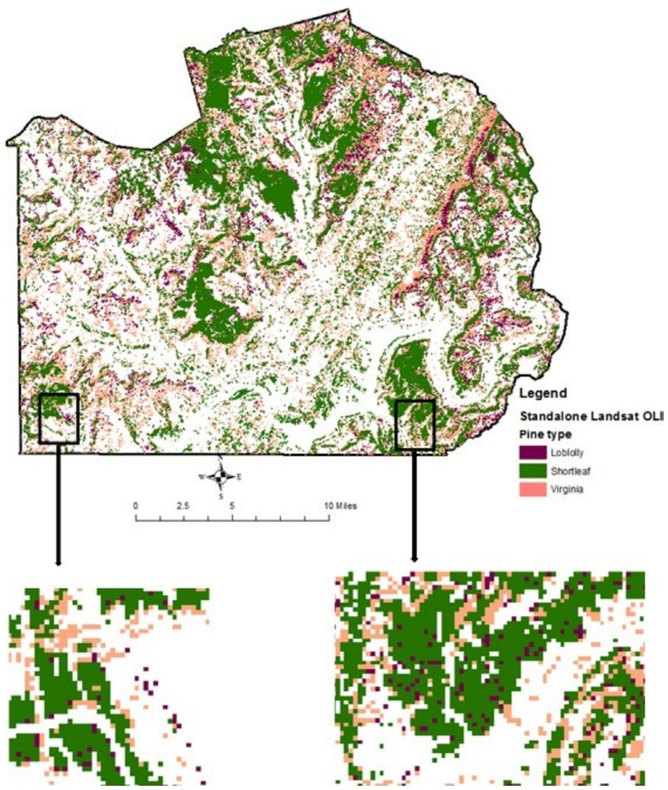


Figure 3. Classification map of southern yellow pines (loblolly, shortleaf, and Virginia pines) derived from stand-alone Landsat OLI satellite data (scenario 1).

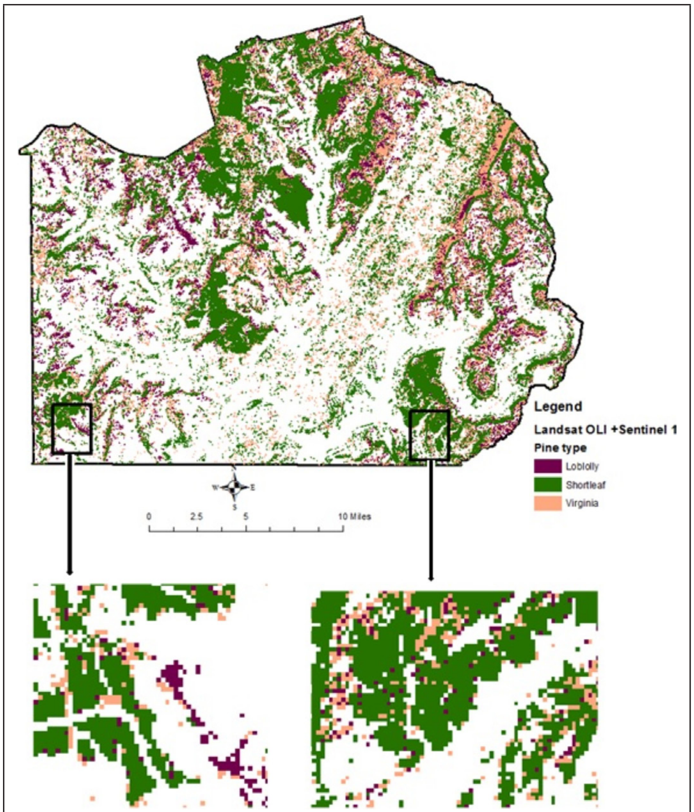


Figure 4. Classification map of southern yellow pines (loblolly, shortleaf, and Virginia pines) derived from the integration of Landsat OLI optical and *Sentinel-1* microwave satellite data (scenario 2).

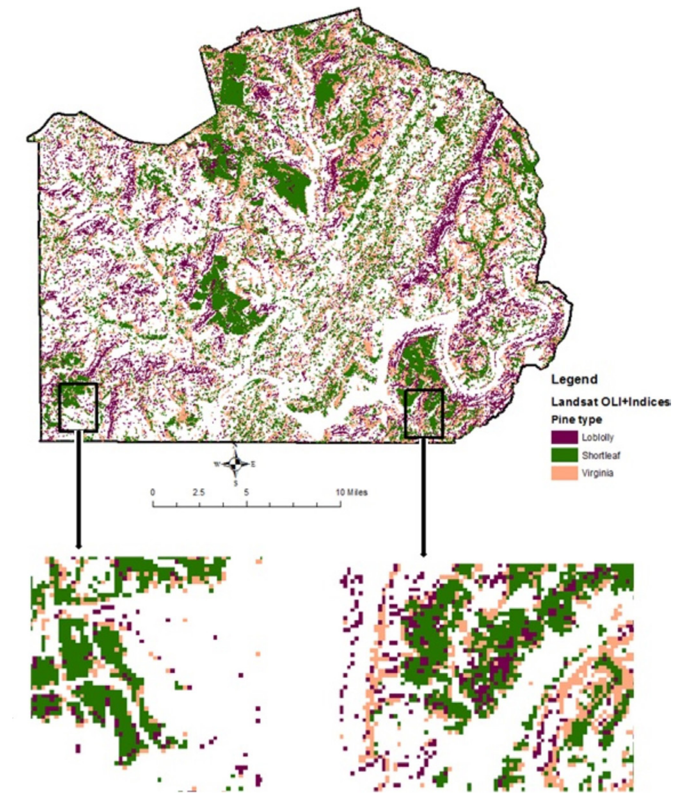


Figure 5. Classification map of southern yellow pines (loblolly, shortleaf, and Virginia pines) derived from the integration of Landsat OLI data and derived vegetation indices (scenario 3).

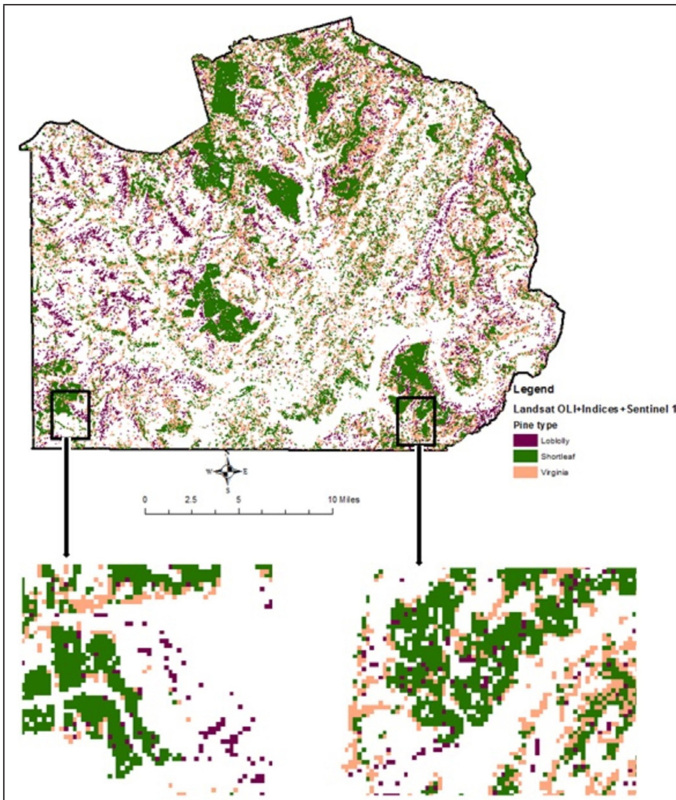


Figure 6. Classification map of southern yellow pines (loblolly, shortleaf, and Virginia pines) derived from the integration of Landsat OLI and *Sentinel-1* data with derived vegetation indices (scenario 4).

Table 4. Classification accuracies of the canopy cover of southern yellow pines derived using four data-integration classification methods.

User Accuracy (%)				
Southern Yellow Pine Class	**Landsat OLI Alone (Scenario 1)**	**Landsat OLI with *Sentinel-1* Data (Scenario 2)**	**Landsat OLI with Derived Vegetation Indices (Scenario 3)**	**Landsat OLI with *Sentinel-1* Data and Derived Vegetation Indices (Scenario 4)**
Loblolly	67	70	87	83
Shortleaf	70	73	75	80
Virginia	63	67	63	67

Producer Accuracy (%)				
Southern Yellow Pine Class	**Landsat OLI Alone (Scenario 1)**	**Landsat OLI with *Sentinel-1* Data (Scenario 2)**	**Landsat OLI with Derived Vegetation Indices (Scenario 3)**	**Landsat OLI with *Sentinel-1* Data and Derived Vegetation Indices (Scenario 4)**
Loblolly	65	66	72	76
Shortleaf	62	66	71	73
Virginia	79	83	86	87

Overall Accuracy (%)				
	Landsat OLI Alone (Scenario 1)	**Landsat OLI with *Sentinel-1* Data (Scenario 2)**	**Landsat OLI with Derived Vegetation Indices (Scenario 3)**	**Landsat OLI with *Sentinel-1* Data and Derived Vegetation Indices (Scenario 4)**
	67	70	75	77

κ Statistics				
	Landsat OLI Alone (Scenario 1)	**Landsat OLI with *Sentinel-1* Data (Scenario 2)**	**Landsat OLI with Derived Vegetation Indices (Scenario 3)**	**Landsat OLI with *Sentinel-1* Data and Derived Vegetation Indices (Scenario 4)**
	0.5	0.54	0.62	0.65

The overall classification accuracy of the canopy cover of southern yellow pines was about 67% when the stand-alone Landsat OLI satellite data set was used (scenario 1; Tables 4 and 5). In this classification method, the user accuracy was highest (70%) for shortleaf pine and lowest (63%) for Virginia pine. In contrast, the producer accuracy was highest (79%) for Virginia pine and lowest (62%) for shortleaf pine (Table 4).

In the classification method in which the Landsat OLI reflectance scene was integrated with *Sentinel-1* backscattering coefficients (scenario 2), the overall accuracy was about 70% compared to reference data (Tables 4 and 6). The overall accuracy increased by about 5% relative to the stand-alone Landsat OLI satellite data. Similarly, other studies have found weaker performance using stand-alone Landsat OLI data in forest canopy cover prediction and mapping compared to integrating Landsat OLI data with *Sentinel-1* microwave data (Poortinga *et al.* 2019; Biswas *et al.* 2020; Li *et al.* 2020). In scenario 2, the user accuracy was highest (73%) for shortleaf and lowest (67%) for Virginia pine. Both loblolly and shortleaf pines had similar producer accuracies, of about 66%, whereas Virginia pine had a producer accuracy of approximately 83% (Table 4). Furthermore, when *Sentinel-1* backscattering coefficients were integrated with Landsat OLI reflectance bands, Virginia pine had a 4% gain in user accuracy, and shortleaf and loblolly pines had a 3% gain. The similar gains in user accuracy imply that the addition of *Sentinel-1* backscattering coefficients is useful for better characterizing loblolly, shortleaf, and Virginia pines. In contrast, shortleaf and Virginia pines had a 4% gain in producer accuracy when *Sentinel-1* backscattering coefficients were integrated, whereas loblolly pine had a 1% gain.

In the classification method in which the Landsat OLI reflectance scene was integrated with satellite-derived vegetation indices (scenario 3), the overall classification accuracy was around 75% compared to reference data (Tables 4 and 7). The overall mapping accuracy of the canopy cover of southern yellow pines increased by about 12% relative to stand-alone Landsat OLI satellite data. Similarly, the results of Matongera *et al.* (2017) also showed that integrating Landsat OLI data with vegetation indices yielded better overall classification accuracy than stand-alone Landsat OLI satellite data. In scenario 3, the user accuracy was highest (87%) for loblolly

Table 5. Error matrix table for the classification of southern yellow pines using stand-alone Landsat OLI satellite data (scenario 1).

Class	**Loblolly**	**Shortleaf**	**Virginia**	**Total**
Reference				
Loblolly	20	8	2	30
Shortleaf	9	28	3	40
Virginia	2	9	19	30
Total	31	45	24	100

Table 6. Error matrix table for the classification of southern yellow pines using integrated Landsat OLI and *Sentinel-1* satellite data (scenario 2).

Class	**Loblolly**	**Shortleaf**	**Virginia**	**Total**
Reference				
Loblolly	21	7	2	30
Shortleaf	9	29	2	40
Virginia	2	8	20	30
Total	32	44	24	100

Table 7. Error matrix table for the classification of southern yellow pines using integrated Landsat OLI data and satellite-derived vegetation indices (scenario 3).

Class	**Loblolly**	**Shortleaf**	**Virginia**	**Total**
Reference				
Loblolly	26	3	1	30
Shortleaf	8	30	2	40
Virginia	2	9	19	30
Total	36	42	22	100

Table 8. Error matrix table for the classification of southern yellow pines using integrated Landsat OLI and *Sentinel-1* satellite data and derived vegetation indices (scenario 4).

Class	**Loblolly**	**Shortleaf**	**Virginia**	**Total**
Reference				
Loblolly	25	4	1	30
Shortleaf	6	32	2	40
Virginia	2	8	20	30
Total	33	44	23	100

and lowest (63%) for Virginia pine. The producer accuracy was highest (86%) for Virginia pine and lowest (71%) for shortleaf pine (Table 4). Furthermore, loblolly pine had the greatest gain in user accuracy (20%), and Virginia pine the least (0%). This implies that the addition of vegetation indices is useful for better characterizing loblolly pine relative to shortleaf and Virginia pines. In contrast, shortleaf pine had the greatest gain in producer accuracy (9%), and loblolly and Virginia pines the least (7%).

In the classification method in which the Landsat OLI reflectance scene was integrated with *Sentinel-1* backscattering coefficients and derived vegetation indices (scenario 4), the overall classification accuracy of southern yellow pines was approximately 77% compared to reference data (Tables 4 and 8). The overall mapping accuracy of the canopy cover of southern yellow pines increased by about 15% compared to stand-alone Landsat OLI satellite data. In scenario 4, the user accuracy was highest (83%) for loblolly and lowest (67%) for Virginia pine. The producer accuracy was highest (87%) for Virginia pine and lowest (73%) for shortleaf pine (Table 4). Furthermore, loblolly pine had the highest gain in user accuracy (16%) in scenario 4 compared to scenario 1, whereas Virginia pine had the lowest (4%). Likewise, both shortleaf and loblolly pines had the highest gain in producer accuracy (11%), and Virginia pine the lowest (8%).

The lower gain in user accuracy for Virginia pine relative to shortleaf and loblolly pines with the addition of *Sentinel-1* backscattering coefficients and derived vegetation indices is possibly due to the morphology of Virginia pine. It has a similar bark color to shortleaf pine—a mix of reddish brown (United States Department of Agriculture 2021)—which possibly increased confusion between Virginia and shortleaf pines in the classification. Consequently, about 27% of Virginia pine was incorrectly classified on the map in scenario 4. Nonetheless, scenario 4 yielded the best overall classification accuracy of the canopy cover of southern yellow pines, whereas the use of stand-alone Landsat OLI data (scenario 1) produced the weakest overall accuracy results in the classification and mapping of the canopy cover of southern yellow pines. Scenario 4 achieved the best overall accuracy because the addition of *Sentinel-1* backscattering coefficients and vegetation indices to Landsat OLI reflectance data improved the spectral resolution and variability of the input variables in the classification. This likely improved the predictive capability of the random-forest classification algorithm. Hence, the addition of backscattering coefficients from *Sentinel-1* and satellite-derived vegetation indices positively contributed to the classification and mapping of the canopy cover of loblolly, shortleaf, and Virginia pines.

Based on the feature-importance score—which estimates which variables were important in the classification process—Landsat OLI spectral band 6 and MSAVI had the highest scores, ranked first and second, respectively. In contrast, IPVI and TSAVI had the lowest scores, ranked fifteenth and sixteenth, respectively. This means that Landsat OLI spectral band 6 and MSAVI were the most important input variables and had high contributions to the classification, whereas IPVI and TSAVI were the least relevant input variables and had low contributions. Therefore, not all the satellite data-derived vegetation indices are necessary in classifying and mapping southern yellow pines using the random-forest classification algorithm. Using just three of the vegetation indices—MSAVI, NDVI, and SAVI—will be enough to improve the classification and mapping of the canopy cover of southern yellow pines. Landsat OLI spectral bands 1 through 5 and 7, the *Sentinel-1* microwave C-band VV, VH, HV, and HH polarizations, NDVI, and SAVI had medium relevance and contributions to the classification, ranked between second and fifteenth based on their feature-importance scores. Therefore, out of the 16 input variables used in the classification process, 14 were relevant and necessary to improve the classification and mapping of southern pines. The use of the random-forest algorithm was better in the data-integration classification methods than the use of other machine-learning algorithms, such as support vector machine, because it provided estimates of the importance of each input variable in the classification process and could be used as a feature-selection tool.

In this study, the 7% decrease in overall classification accuracy of southern yellow pines produced by integrating Landsat OLI data with *Sentinel-1* backscattering coefficients compared to using vegetation indices was not expected. This implies that vegetation indices could contribute more to the classification and mapping of the canopy cover of southern yellow pines than *Sentinel-1* backscattering coefficients. However, to attain the best prediction and mapping of the canopy cover of loblolly, shortleaf, and Virginia pines, the integration of the Landsat OLI reflectance scene with *Sentinel-1* backscattering coefficients and derived vegetation indices is relevant.

Future research will examine how other machine-learning classification algorithms, such as gradient-boosted tree, extreme gradient boosting, and multi-layer perceptron, perform against the random-forest classifier in mapping southern yellow pines using the Landsat OLI reflectance scene with *Sentinel-1* backscattering coefficients and derived vegetation indices. Furthermore, exploring the integration of Landsat OLI optical data with *Sentinel-1* C-band SAR sensor and lidar data in other natural-resources applications, such as wetlands and agriculture, is an area of further research.

Conclusion

This study successfully examined the integration of Landsat OLI optical data with *Sentinel-1* microwave satellite data and derived vegetation indices in mapping the canopy cover of loblolly, shortleaf, and Virginia pines. We found that when Landsat OLI data was integrated with *Sentinel-1* backscattering coefficients, the classification of the canopy cover of southern yellow pines increased by about 5% compared to stand-alone Landsat OLI satellite data. Similarly, the integration of Landsat OLI reflectance bands with satellite data-derived vegetation indices increased the overall mapping accuracy by about 12% compared to stand-alone Landsat OLI satellite data. Furthermore, the best overall classification accuracy (77%) of the canopy cover of southern yellow pines was produced when the Landsat OLI reflectance scene was integrated with *Sentinel-1* backscattering coefficients and derived vegetation indices. Landsat OLI spectral band 6 and MSAVI were the most important input variables in the classification of the canopy cover, and IPVI and TSAVI were the least important variables. The classification method that integrated Landsat OLI optical data with *Sentinel-1* microwave satellite data and derived vegetation indices can be easily developed to successfully map the canopy cover of southern yellow pines.

Acknowledgments

Many thanks to Jim D. Lane at Prentice Cooper State Forest and Brian Hughett at the Tennessee Department of Agriculture, Forestry Section, for providing the field-location data set used to carry out this project. Our appreciation goes to the United States Department of Agriculture for providing support through the McIntire-Stennis funding program.

References

Abdollahnejad, A., D. Panagiotidis and L. Bílek. 2019. An integrated GIS and remote sensing approach for monitoring harvested areas from very high-resolution, low-cost satellite images. *Remote Sensing* 11(21):2539.

Akumu, C. E., E. O. Amadi and S. Dennis. 2021. Application of drone and WorldView-4 satellite data in mapping and monitoring grazing land cover and pasture quality: Pre- and post-flooding. *Land* 10(3):321.

Akumu, C. E., J. Henry, T. S. Gala, S. Dennis, C. Reddy, F. Tegegne, S. Haile and R. S. Archer. 2018. Inland wetlands mapping and vulnerability assessment using an integrated geographic information system and remote sensing techniques. *Global Journal of Environmental Science and Management* 4(4):387–400.

Baret, F., G. Guyot and D. J. Major. 1989. TSAVI: A vegetation index which minimizes soil brightness effects on LAI and APAR estimation. Pages 1355–1358 in *12th Canadian Symposium on Remote Sensing Geoscience and Remote Sensing Symposium*, held in Vancouver, British Columbia, 10–14 July 1989. Edited by J. Editor. Location: IEEE.

Bera, B., S. Saha and S. Bhattacharjee. 2020. Estimation of forest canopy cover and forest fragmentation mapping using Landsat satellite data of Silabati River Basin (India). *KN—Journal of Cartography and Geographic Information* 70:181–197.

Biswas, S., Q. Huang, A. Anand, M. S. Mon, F.-E. Arnold and P. Leimgruber. 2020. A multi sensor approach to forest type mapping for advancing monitoring of Sustainable Development Goals (SDG) in Myanmar. *Remote Sensing* 12(19):3220.

Clabo, D. C. and W. Clatterbuck, eds. 2005. *A Tennessee Landowner and Practitioner Guide for Establishment and Management of Shortleaf and Other Pines. Publication PB1751*. Location: University of Tennessee Extension, Institute of Agriculture.

Costa, H., D. Almeida, F. Vala, F. Marcelino and M. Caetano. 2018. Land cover mapping from remotely sensed and auxiliary data for harmonized official statistics. *ISPRS International Journal of Geo-Information* 7(4):157. https://doi.org/10.3390/ijgi7040157.

Crippen, R. E. 1990. Calculating the vegetation index faster. *Remote Sensing of Environment* 34(1):71–73.

Deng, S., Y. Shi, Y. Jin and L. Wang. 2011. A GIS-based approach for quantifying and mapping carbon sink and stock values of forest ecosystem: A case study. *Energy Procedia* 5:1535–1545.

Elmahdy, S. I., T. A. Ali, M. M. Mohamed, F. M. Howari, M. Abouleish and D. Simonet. 2020. Spatiotemporal mapping and monitoring of mangrove forests changes from 1990 to 2019 in the Northern Emirates, UAE using random forest, kernel logistic regression and naive Bayes tree models. *Frontiers in Environmental Science* 8:102. https://doi.org/10.3389/fenvs.2020.00102.

Elmaz, F., B. Büyüçakır, Ö. Yücel and A. Y. Mutlu. 2020. Classification of solid fuels with machine learning. *Fuel* 266:117066. https://doi.org/10.1016/j.fuel.2020.117066.

English, B., J. Menard and K. Jensen. 2004. *Tennessee's Forest and Forest Products Industry and Associated Economic Impacts for 2000. Research Series 01-04*. Knoxville, Tenn.: The University of Tennessee, Agricultural Experiment Station.

Fatoyinbo, T. E. and A. H. Armstrong. 2010. Remote characterization of biomass measurements: Case study of mangrove forests. In *Biomass*, edited by M. Momba and F. Bux, 65–78. Rjeka, Croatia: InTech.

Ganz, S., P. Adler and G. Kändler. 2020. Forest cover mapping based on a combination of aerial images and Sentinel-2 satellite data compared to National Forest Inventory data. *Forests* 11:1322.

Gao, B.-C. 1996. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment* 58(3):257–266.

Hamilton, D. A., K. L. Brothers, S. D. Jones, J. Colwell and J. Winters. 2021. Wildland fire tree mortality mapping from hyperspatial imagery using machine learning. *Remote Sensing.* 13(2):290.

Hansen, M. C., A. Egorov, P. V. Potapov, S. V. Stehman, A. Tyukavina, S. A. Turubanova, D. P. Roy, S. J. Goetz, T. R. Loveland, J. Ju, A. Kommareddy, V. Kovalskyy, C. Forsyth and T. Bents. 2014. Monitoring conterminous United States (CONUS) land cover change with Web-Enabled Landsat Data (WELD). *Remote Sensing of the Environment* 140:466–484.

Hinkle, C. R. 1989. Forest communities of the Cumberland Plateau of Tennessee. *Journal of the Tennessee Academy of Science* 64(3):123–129.

Hodges, J. A., R. J. Norell and M. H. Sarah. 2018. Tennessee. <www.britannica.com/place/Tennessee> Accessed 25 December 2020.

Huete, A. R. 1988. A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment* 25(3):295–309.

Jiménez, E., J. A. Vega, J. M. Fernández-Alonso, D. Vega-Nieva, L. Ortiz, P. M. López-Serrano and C. A. López-Sánchez. 2017. Estimation of aboveground forest biomass in Galicia (NW Spain) by the combined use of LiDAR, LANDSAT ETM+ and National Forest Inventory data. *iForest—Biogeosciences and Forestry* 10(3):590–596.

Ke, Y., L. J. Quackenbush and J. Im. 2010. Synergistic use of QuickBird multispectral imagery and LIDAR data for object-based forest species classification. *Remote Sensing of Environment* 114(6):1141–1154.

Li, W., Z. Niu, R. Shang, Y. Qin, L. Wang and H. Chen. 2020. High-resolution mapping of forest canopy height using machine learning by coupling ICESat-2 LiDAR with Sentinel-1, Sentinel-2 and Landsat-8 data. *International Journal of Applied Earth Observation and Geoinformation* 92:102163.

Lichtenthaler, H. K., M. Lang, M. Sowinska, F. Heisel and J. A. Miehé. 1996. Detection of vegetation stress via a new high resolution fluorescence imaging system. *Journal of Plant Physiology* 148(5):599–612.

Martinuzzi, S., W. A. Gould, O. M. Ramos Gonzalez, A. M. Robles, P. C. Maldonado, N. Pérez-Buitrago and J. J. Fumero Caban. 2008. Mapping tropical dry forest habitats integrating Landsat NDVI, Ikonos imagery, and topographic information in the Caribbean island of Mona. *Revista de Biología Tropical* 56(2):625–639.

Mather, P. M. and M. Koch. 2011. *Computer Processing of Remotely-Sensed Images: An Introduction*, 4th ed. Chichester, United Kingdom: Wiley-Blackwell.

Matongera, T. N., O. Mutanga, T. Dube and M. Sibanda. 2017. Detection and mapping the spatial distribution of bracken fern weeds using the Landsat 8 OLI new generation sensor. *International Journal of Applied Earth Observation and Geoinformation* 57:93–103.

Nguyen, T. 2015. Optimal ground control points for geometric correction using genetic algorithm with global accuracy. *European Journal of Remote Sensing* 48(1):101–120.

Pehani, P., K. Čotar, A. Marsetič, J. Zaletelj and K. Oštir. 2016. Automatic geometric processing for very high resolution optical satellite data based on vector roads and orthophotos. *Remote Sensing* 8(4):343. https://doi.org/10.3390/rs8040343.

Poortinga, A., K. Tenneson, A. Shapiro, Q. Nquyen, K. San Aung, F. Chishtie and D. Saah. 2019. Mapping plantations in Myanmar by fusing Landsat-8, Sentinel-2 and Sentinel-1 data along with systematic error quantification. *Remote Sensing* 11(7):831.

Prabhakara, K., W. D. Hively and G. W. McCarty. 2015. Evaluating the relationship between biomass, percent groundcover and remote sensing indices across six winter cover crop fields in Maryland, United States. *International Journal of Applied Earth Observation and Geoinformation* 39:88–102.

Qi, J., A. Chehbouni, A. R. Huete, Y. H. Kerr and S. Sorooshian. 1994. A modified soil adjusted vegetation index. *Remote Sensing of Environment* 48(2):119–126.

Reid, S. L., J. L. Walker and A. Schaaf. 2016. Using multi-spectral Landsat imagery to examine forest health trends at Fort Benning, Georgia. Pages 600–603 in *Proceedings of the 18th Biennial Southern Silvicultural Research Conference*, held in Knoxville, Tennessee, March 2-5, 2015. Edited by C.J. Schweitzer, W.K.

Rhyma, P. P., K. Norizaha, O. Hamdan, I. Faridah-Hanum and A. W. Zulfa. 2020. Integration of normalised different vegetation index and Soil-Adjusted Vegetation Index for mangrove vegetation delineation. *Remote Sensing Applications: Society and Environment* 17:100280.

Roth, K. L., D. A. Roberts, P. E. Dennison, S. H. Peterson and M. Alonzo. 2015. The impact of spatial resolution on the classification of plant species and functional types within imaging spectrometer data. *Remote Sensing of Environment* 171:45–57.

Rozenstein, O. and A. Karnieli. 2011. Comparison of methods for land-use classification incorporating remote sensing and GIS inputs. *Applied Geography* 31(2):533–544.

Shang, X. and L. A. Chisholm. 2014. Classification of Australian native forest species using hyperspectral remote sensing and machine-learning classification algorithms. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7(6):2481–2489.

Sharma, R. C., K. Hara and H. Hirayama. 2017. A machine learning and cross-validation approach for the discrimination of vegetation physiognomic types using satellite based multispectral and multitemporal data. *Scientifica* 2017:9086479. https://doi.org/10.1155/2017/9806479.

Sinha, S., L. K. Sharma and M. S. Nathawat. 2015. Improved Land-use/Land-cover classification of semi-arid deciduous forest landscape using thermal remote sensing. *The Egyptian Journal of Remote Sensing and Space Sciences* 18(2):217–233.

Sjöqvist, H., M. Längqvist and F. Javed. 2020. An analysis of fast learning methods for classifying forest cover types. *Applied Artificial Intelligence* 34(10):691–709.

Skidmore, A. K., W. Bijker, K. Schmidt and L. Kumar. 1997. Use of remote sensing and GIS for sustainable land management. *ITC Journal*, 3/4:302–315.

Starnes, D. D. 1986. *Tennessee Topology.* Location: Tennessee Department of Environment and Conservation.

Tawfeik, M., H. Elhifnawy, E. Hamza and A. Shawky. 2016. Determination of suitable requirements for geometric correction of remote sensing satellite images when using ground control points. *International Research Journal of Engineering and Technology* 3(10):54–62.

Twele, A., W. Cao, S. Plank and S. Martinis. 2016. Sentinel-1-based flood mapping: A fully automated processing chain. *International Journal of Remote Sensing* 37(13):2990–3004.

United States Department of Agriculture. 2021. Fire Effects Information System (FEIS) Index of Species Information. <https://www.fs.fed.us/database/feis/plants/tree/pinech/all.html> Accessed 12 May 2021.

United States Geological Survey. 2019. *Landsat 8 (L8) Data Users Handbook*, version 5.0. Location: United States Geological Society.

van Aardt, J.A.N. and R. H. Wynne. 2007. Examining pine spectral separability using hyperspectral data from an airborne sensor: Extension of field-based results. *International Journal of Remote Sensing* 28(2):431–436.

van Ewijk, K. Y., C. F. Randin, P. M. Treitz and N. A. Scott. 2014. Predicting fine-scale tree species abundance patterns using biotic variables derived from LiDAR and high spatial resolution imagery. *Remote Sensing of Environment* 150:120–131.

Vincent, S., S.A.J. Francis, K. Raimond and O. P. Kumar. 2019. A study of types of sensors used in remote sensing. *International Journal of Electronics and Telecommunications* 65(2):217–228.

Wan, H., Y. Tang, L. Jing, H. Li, F. Qiu and W. Wu. 2021. Tree species classification of forest stands using multisource remote sensing data. *Remote Sensing* 13(1):144.

Xie, Y., Z. Sha and M. Yu. 2008. Remote sensing imagery in vegetation mapping: A review. *Journal of Plant Ecology* 1(1):9–23.

Xu, H. 2006. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing* 27(14):3025–3033.

Xue, J. and B. Su. 2017. Significant remote sensing vegetation indices: A review of developments and applications. *Journal of Sensors* 2017:1353691.

Young, T. M., D. G. Hodges and T. G. Rials. 2007. The forest products economy of Tennessee. *Forest Products Journal* 57(4):12–19.

Augmented Sample-Based Real-Time Spatiotemporal Spectral Unmixing

Xinyu Ding and Qunming Wang

Abstract

Recently, the method of spatiotemporal spectral unmixing (STSU) was developed to fully explore multi-scale temporal information (e.g., MODIS–Landsat image pairs) for spectral unmixing of coarse time series (e.g., MODIS data). To further enhance the application for timely monitoring, the real-time STSU (RSTSU) method was developed for real-time data. In RSTSU, we usually choose a spatially complete MODIS–Landsat image pair as auxiliary data. Due to cloud contamination, the temporal distance between the required effective auxiliary data and the real-time data to be unmixed can be large, causing great land cover changes and uncertainty in the extracted unchanged pixels (i.e., training samples). In this article, to extract more reliable training samples, we propose choosing the auxiliary MODIS–Landsat data temporally closest to the prediction time. To deal with the cloud contamination in the auxiliary data, we propose an augmented sample-based RSTSU (ARSTSU) method. ARSTSU selects and augments the training samples extracted from the valid (i.e., non-cloud) area to synthesize more training samples, and then trains an effective learning model to predict the proportions. ARSTSU was validated using two MODIS data sets in the experiments. ARSTSU expands the applicability of RSTSU by solving the problem of cloud contamination in temporal neighbors in actual situations.

Introduction

Mixed pixels have been considered one of the main factors restricting the reliability of land cover mapping in remote sensing (Schowengerdt 1997; Lin *et al.* 2019; Wu *et al.* 2019, 2021). For more reliable interpretation of mixed pixels, scholars have increasingly devoted attention to spectral unmixing (also known as soft classification; Settle and Drake 1993). Spectral unmixing can predict the proportions of land cover classes in mixed pixels and can provide more detailed land cover information than traditional hard classification (Foody 2002; Atkinson 2005). In recent years, a series of methods for spectral unmixing have been developed. It is beyond the scope of this article to review these works, but readers can refer to several reviews (Keshava and Mustard 2002; Quintano *et al.* 2012; Heylen *et al.* 2014; Shi and Wang 2014; Bhatt and Joshi 2020).

The main factors affecting the reliability of spectral unmixing include the extraction of pure end members (i.e., the representative spectrum of each land cover class) and the intraclass spectral variation (i.e., the same land cover class presenting different spectra; Zhang *et al.* 2019). Although a number of multiple end member-based methods have been proposed for the problem of intraclass spectral variation (Roberts *et al.* 1998; Bateson *et al.* 2000), the extraction of a large number of end members is still a huge challenge, especially in areas with great heterogeneity. Spectral-unmixing methods based on machine learning do not require pure end members, but they still need a great deal of supervised information (i.e., the proportion of each land cover class in the mixed pixels) to train learning models. Moreover, most of the spectral-unmixing methods are carried out using images on a single day, and the few studies on handling time-series data are not suitable for dynamic monitoring (Zurita-Milla *et al.* 2011; Deng and Zhu 2020). To enhance the reliability of spectral unmixing, in our previous work (Q. Wang *et al.* 2021a), we proposed a spatiotemporal spectral-unmixing (STSU) method for spectral unmixing of MODIS time series, which explores the multi-scale spatiotemporal information in auxiliary MODIS–Landsat pairs. STSU detects land cover change and extracts supervised information of mixed spectra in MODIS images at the prediction time, so as to construct training samples to train learning models. The STSU model needs MODIS–Landsat data both before and after the prediction time, which is suitable for handling historical time-series data. For areas where the land cover changes rapidly, it is necessary to exploit real-time images for timely monitoring—that is, it is of important application value to investigate spectral unmixing of real-time data.

In response to these requirements, Q. Wang *et al.* (2021b) proposed a real-time spatiotemporal spectral-unmixing (RSTSU) method. Different from STSU, RSTSU was designed for real-time data, and needs only the coarse–fine spatial-resolution image pair before the real-time data. Moreover, RSTSU performs change detection on two coarse-spatial-resolution images to extract unchanged pixels at the prediction time as training samples, and trains an effective machine-learning model. Based on the trained model, the proportion of all pixels in the real-time coarse image can be predicted.

Under normal circumstances, in the RSTSU method, spatially complete (e.g., without cloud contamination) auxiliary data at the previous time are required. Therefore, the time interval between the effective auxiliary data and the prediction data is usually longer than expected, and the land cover in the two images may change greatly. Correspondingly, there can be greater uncertainty in the unchanged pixels extracted by change detection. Thus, to enhance the reliability of training samples, we should choose the data temporally closest to the prediction time. However, the image at the closest time is susceptible to cloud contamination (Chen *et al.* 2017; Q. Wang *et al.* 2020; Q. Wang, Wang *et al.* 2021), leading to missing data in part of the image. Meanwhile, due to the reduction in the number of effective auxiliary data, the number of extracted training samples is also reduced, imposing a negative effect on the training process and thereby affecting the final spectral-unmixing predictions. It should be noted that a number of methods have been developed for cloud removal, but most of them need temporally close data with spatially complete coverage (Shen *et al.* 2015; Gao and Gu 2017; Goward *et al.* 2019).

Xinyu Ding and Qunming Wang (corresponding author) are with the College of Surveying and Geo-Informatics, Tongji University, Shanghai, China (wqm11111@126.com).

Contributed by Alper Yilmaz, July 1, 2021 (sent for review July 14, 2021; reviewed by Ming Hao, Zhongbin Li).

Photogrammetric Engineering & *Remote Sensing*
Vol. 88, No. 1, January 2022, pp. 39–45.
0099-1112/22/39–45

doi: 10.14358/PERS.21-00039R2

Based on the influence of cloud contamination in the auxiliary data, we propose an augmented sample-based real-time spatiotemporal spectral-unmixing (ARSTSU) model. Similarly, we consider spectral unmixing of MODIS data (Q. Wang, Peng *et al.* 2021). ARSTSU makes full use of the cloud-free data in the auxiliary Landsat and MODIS data at the closest time. It constructs training samples based on change detection between the common cloud-free MODIS data at two times. Then the most reliable samples are selected from the original training samples, which are further augmented to generate more training samples. The ARSTSU model inherits all the advantages of RSTSU, which does not require end-member extraction and can consider intraclass spectral variation. Meanwhile, it is suitable for spectral unmixing of real-time data and dynamic monitoring of land cover changes. Unlike RSTSU, however, ARSTSU can use the auxiliary data (i.e., the MODIS–Landsat image pair) temporally closest to the real-time data. Thus, ARSTSU expands the applicability of the RSTSU model by coping with the problem of cloud contamination in actual situations.

The remainder of this article is organized as follows. In the next section, the principle of the ARSTSU method is introduced explicitly. The following section demonstrates the effectiveness of ARSTSU through experiments on data sets covering two regions and comparing it with three benchmark methods. Then the conclusion is drawn.

Methods

The flowchart of ARSTSU—training-sample extraction, selection and augmentation, model training, and prediction—is shown in Figure 1. The principle of each part is introduced in the following.

Extraction of Training Samples (Based on Auxiliary Data with Cloud Contamination)

Suppose we need to predict the proportion for MODIS data at T_n, and the temporally closest MODIS and Landsat data are at T_m (both covered by cloud). First, for the cloud-free region, we calculate the spectral difference between MODIS pixels at T_m and T_n. Then we calculate the modulus value of the spectral difference for each pixel to obtain the modulus image. A smaller modulus value means that the land cover within that pixel it is less likely to be changed. Finally, based on the modulus image, the OTSU algorithm is used to segment the image into two classes to realize change detection. That is, a pixel with a difference less than the OTSU-determined threshold is identified as an unchanged pixel, and its spatial location is recorded as $\mathbf{x}$.

For each unchanged pixel, the proportion set of the land cover classes at T_n is equal to that at T_m—that is, $\mathbf{P}(T_n, \mathbf{x}) = \mathbf{P}(T_m, \mathbf{x})$. For the latter, the Landsat data at T_m are used for hard classification, and then the classification map is upscaled to the MODIS spatial resolution to obtain the proportion of each land cover class at T_m:

$$P_k\left(T_m, \mathbf{x}\right) = \sum_{f=1}^{z^2} I_k\left(T_m, \mathbf{x}(f)\right) / z^2, \quad k = 1, 2, \ldots, N_c \tag{1}$$

where $\mathbf{x}$ is the spatial location of an unchanged pixel at the MODIS spatial resolution, $\mathbf{x}(f)$ is the unchanged pixel at the Landsat spatial resolution falling within $\mathbf{x}$, N_c is the number of land cover classes, z is the scale factor between the MODIS and Landsat data, $P_k(T_m, \mathbf{x})$ is the proportion of the kth class for the MODIS pixel at $\mathbf{x}$ spatially and at T_m temporally, and $I_k(T_m, \mathbf{x}(f))$ is the class indicator of the kth class for the Landsat pixel at $\mathbf{x}(f)$ spatially and at T_m temporally (it equals 1 if the Landsat pixel belongs to the k^{th} class, and 0 otherwise).

Moreover, based on the spectra of the unchanged MODIS pixels at T_n—that is, $\mathbf{M}(T_n, \mathbf{x})$—and the corresponding proportion $\mathbf{P}(T_n, \mathbf{x})$, a training-sample set can be constructed, which is denoted as $\mathbf{S}_1$. Meanwhile, the spectra of all remaining pixels (i.e., changed pixels) in the MODIS data at T_n are used as the prediction samples, and the spatial location of a changed pixel is recorded as $\mathbf{y}$.

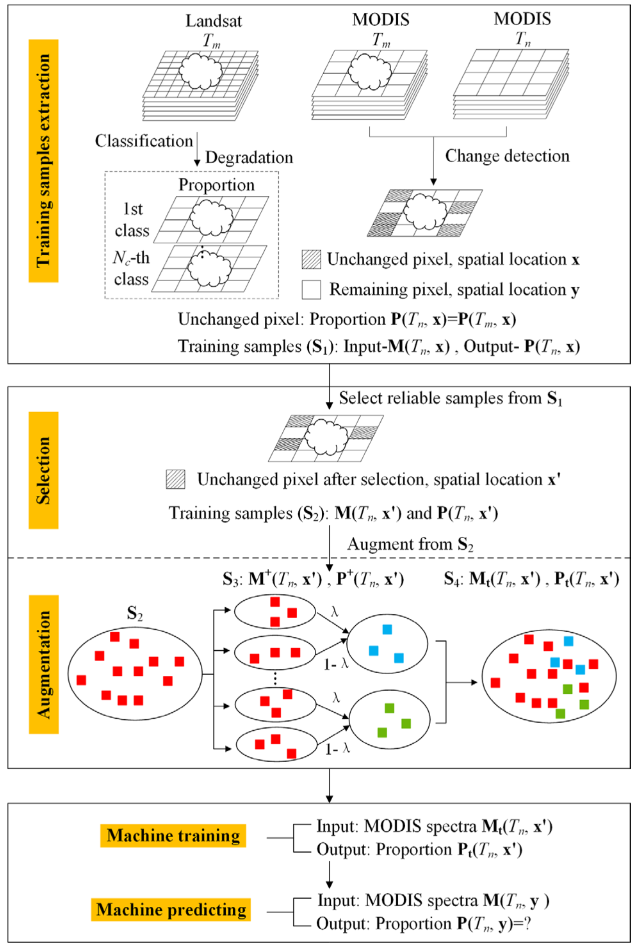


Figure 1. Flowchart of the augmented sample-based real-time spatiotemporal spectral-unmixing method.

Selection and Augmentation of Training Samples

As the auxiliary data at T_m are affected by cloud contamination, only part of the data can be used for change detection, leading to a limited number of unchanged pixels detected between T_m and T_n. This issue is prominent when the cloud size is large. Simultaneously, the training samples extracted by the OTSU algorithm contain unavoidable uncertainty, especially for samples with large modulus values. To ensure a sufficient number of reliable training samples, we sequentially select and augment the training samples obtained by the previous change detection. Specifically, in the original training-sample set $\mathbf{S}_1$, a certain percentage of the pixels with the smallest modulus value are selected as reliable samples, and the corresponding pixel locations are recorded as $\mathbf{x}'$. The filtered sample set is recorded as $\mathbf{S}_2$.

For sample augmentation, the training-sample set $\mathbf{S}_2$ is randomly divided into several groups, with same number of samples for each group. Then we randomly select the pairwise samples from two groups, and the corresponding spectra and proportions are linearly mixed simultaneously:

$$\begin{aligned} \mathbf{M}^+(T_n, \mathbf{x}') &= \lambda \mathbf{M}(T_n, \mathbf{x}_i')+(1-\lambda)\mathbf{M}(T_n, \mathbf{x}_j') \\ \mathbf{P}^+(T_n, \mathbf{x}') &= \lambda \mathbf{P}(T_n, \mathbf{x}_i')+(1-\lambda)\mathbf{P}(T_n, \mathbf{x}_j') \end{aligned} \tag{2}$$

where $\mathbf{M}(T_n, \mathbf{x}_i')$ and $\mathbf{M}(T_n, \mathbf{x}_j')$ are the spectra of training samples at $\mathbf{x}_i'$ and $\mathbf{x}_j'$ (from two different groups), $\mathbf{P}(T_n, \mathbf{x}_i')$ and $\mathbf{P}(T_n, \mathbf{x}_j')$ are the proportions for the corresponding samples, $\lambda \in (0,1)$ is a linear mixing coefficient, and $\mathbf{M}^{+}(T_n, \mathbf{x}')$ and $\mathbf{P}^{+}(T_n, \mathbf{x}')$ are the MODIS spectra and the corresponding proportion for the augmented sample. The augmented-sample set is denoted as $\mathbf{S}_3$. It should be noted that noise can also be considered during the augmentation process. In this article, to simplify the process, the influence of noise is ignored.

When the mixing coefficient λ varies, different augmented samples will be generated. In this article, we randomly divide the sample set $\mathbf{S}_2$ into 10 groups, and mix the samples in the first and 10th groups with coefficients of 0.1 and 0.9, respectively. Similarly, the samples in the second and the ninth group are mixed with coefficients of 0.2 to 0.8, and the remaining groups are mixed in the same manner.

Finally, we combine the training-sample sets $\mathbf{S}_2$ and $\mathbf{S}_3$ to form a new training-sample set $\mathbf{S}_4$:

$$\mathbf{S}_4 = \mathbf{S}_2 \cup \mathbf{S}_3 \tag{3}$$

The training-sample set $\mathbf{S}_4$ is used subsequently for training a learning model. The spectrum in the training-sample set $\mathbf{S}_4$ is denoted as $\mathbf{M}_t(T_n, \mathbf{x}_i')$, and the corresponding proportion is denoted as $\mathbf{P}_t(T_n, \mathbf{x}_i')$.

Machine Training and Prediction (for Remaining Samples)

In this article, we consider the least-squares support vector machine (LSSVM; Suykens *et al.* 2002; H. Wang *et al.* 2014) as the machine-learning model. LSSVM can cope with pattern classification and regression problems, which turns the inequality constraints in a support vector machine (Vapnik 1995; Pal and Foody 2010) into equality constraints, and simplifies the calculation process. Based on the training-sample set $\mathbf{S}_4$, the LSSVM learning model is trained to predict the fitting parameters characterizing the relation between proportion $\mathbf{P}_t(T_n, \mathbf{x}')$ and pixel spectrum $\mathbf{M}_t(T_n, \mathbf{x}')$. After the training model is acquired, the spectrum of the remaining MODIS pixels at T_n—that is, $\mathbf{M}(T_n, \mathbf{y})$—is used as the input of the prediction sample, and the proportion $\mathbf{P}(T_n, \mathbf{y})$ of the sample can be predicted.

Experiments

Data Sets and Experimental Setup

We selected MODIS and Landsat data sets from two different regions in China (Wuhan, Hubei, and Daxing, Beijing) for experimental validation. In each region, the 480-m MODIS and 30-m Landsat image pairs at two time points were selected. Both types of data contain six bands (blue, green, red, near-infrared, short-wave infrared 1, and short-wave infrared 2). For the Wuhan data, the acquisition dates were 3 May 2001 and 22 July 2001; the spatial sizes of the MODIS data (from the MOD09GA product) and the Landsat data were, respectively, 62×62 and 992×992 pixels. For the Daxing data, the acquisition dates were 4 September 2014 and 6 October 2014, and the spatial sizes of the MODIS and Landsat data were, respectively, 60×60 and 720×720 pixels. The MODIS data are the MOD02HKM product that has been atmospherically corrected (from Li *et al.* 2020). For the two Landsat images in each region, we performed unsupervised classification (coupled with visual interpretation) to produce 30-m vegetation and non-vegetation maps at the two time points. With regard to the vegetation map, we upscaled it to 480-m MODIS spatial resolution to produce the reference image of the vegetation proportion at the corresponding time.

For each region, the MODIS and Landsat data at the previous time were used as temporally neighboring auxiliary data for spectral unmixing of the MODIS data at the latter time. Then we simulated cloud contamination for the data at the previous time. In the following sections, we consider a cloud with ideal square shape covering 40% of the entire data and then use the cloud masks from three real MODIS images acquired on other days (covering the spatially close area in the Wuhan region). Finally, the Landsat-derived proportion at the latter time (that is, the prediction time) is used as reference data for evaluating accuracy. Figure 2 shows the MODIS image and the corresponding proportion reference image at the prediction time for each region. The simulated temporally neighboring MODIS–Landsat data with cloud contamination are shown in Figures 3 and 4.

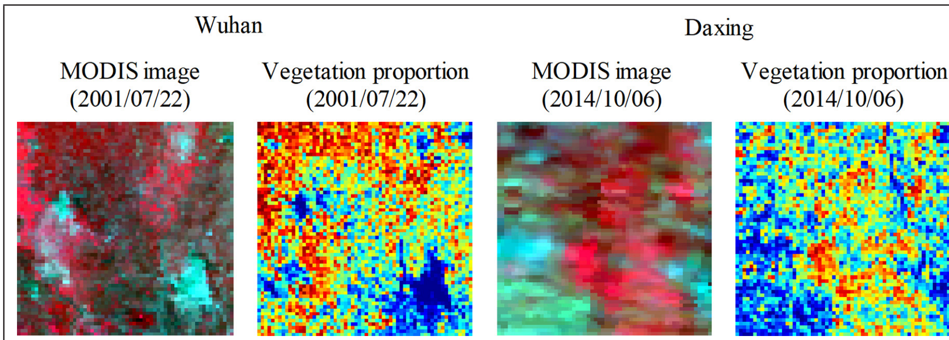


Figure 2. MODIS images (near-infrared, red, and green shown as red, green, and blue channels) and vegetation-proportion images at the prediction time.

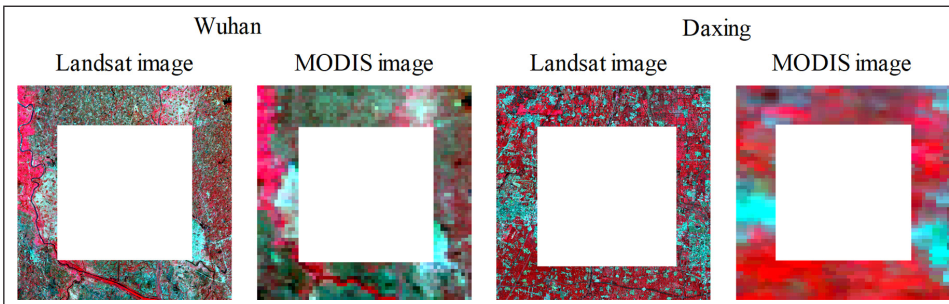


Figure 3. The temporally neighboring MODIS–Landsat images with simulated cloud cover (40% of the entire region; white).

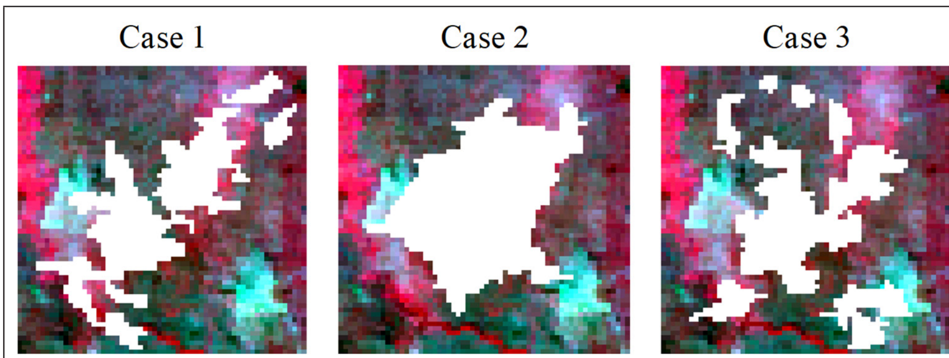


Figure 4. The temporally neighboring MODIS images simulated with cloud masks obtained from real MODIS images for the Wuhan region.

In the experiments, we chose three benchmark methods:

LSMM (linear spectral mixture model). If the extracted unchanged pixel is a pure pixel at T_m, it will also be recognized as a pure pixel at T_n. In the two study regions, completely pure pixels are rarely or almost never available. Therefore, if the proportion of vegetation or non-vegetation within a pixel is larger than 90%, we considered it a pure pixel. Moreover, the average of the spectra of all the pure pixels for each class was used as the end member in the linear spectral mixture model to predict the proportions.

LSSVM. The spectra of all MODIS pixels without cloud cover at T_m and the corresponding Landsat-derived proportions were used to construct training samples and to train a LSSVM. The proportions of all MODIS pixels at T_n were predicted based on the trained LSSVM model.

RSTSU. The difference between ARSTSU and RSTSU is that the latter uses only the training samples extracted from the cloud-free area—that is, $\mathbf{S}_1$ in Figure 1.

Results of the Data Simulated with Square Cloud

Extraction of Training Samples

For the MODIS difference images in the two experimental regions, the OTSU algorithm was used to detect unchanged pixels in the cloud-free area; the results are shown in Figure 5. It can be seen that the number of unchanged pixels (in black) extracted from the cloud-free area is relatively limited compared to the entire image, indicating the necessity of sample augmentation.

Comparison Between Different Methods

Correlation coefficient (CCs), root-mean-square error (RMSE), and mean absolute error (MAE) were used to evaluate accuracy. Figure 6 shows the accuracy evaluation of the spectral-unmixing results for the four models; the *x*-coordinate represents the percentage of reliable samples selected by the ARSTSU model. It can be clearly seen that for the two experimental regions, the CCs of the ARSTSU results are larger than those of the LSMM, LSSVM, and RSTSU results for various percentages of reliable samples, and the RMSEs and MAEs of the ARSTSU results are also smaller. More precisely, for the Wuhan region, when the percentage of reliable samples is 40%, the CC of ARSTSU is 0.07, 0.09, and 0.07 larger than LSMM, LSSVM, and RSTSU, respectively. Correspondingly, the RMSE of ARSTSU is 0.10, 0.23, and 0.02 smaller. Similarly, for the Daxing region, when the percentage of reliable samples is 40%, the CCs of ARSTSU are 0.20, 0.04, and 0.02 larger than LSMM, LSSVM, and RSTSU, and the MAEs are 0.25, 0.10, and 0.01 smaller. In addition, when the percentage of reliable samples decreases, the accuracy of ARSTSU increases gradually. When the former drops to 60%, the latter stabilizes.

Figure 7 shows the proportion maps for the four spectral-unmixing methods. The error maps (in absolute values) are shown in Figure 8, and the scatterplots of vegetation proportion (prediction versus reference) are shown in Figure 9. We can draw the following conclusions. First, because the LSMM model cannot take the intraclass spectral variation into account, large errors occur for regions with great heterogeneity, resulting in spectral-unmixing results that present a large cluster of blocks. Second, due to changes in land cover and imaging conditions, the training data extracted directly from a temporally neighboring image pair are quite different from the data at prediction time, leading to results with obvious errors. For example, in Figure 7, the LSSVM result in Wuhan is dominated by red pixels (i.e., almost covered by vegetation),

Figure 5. Segmentation results based on MODIS difference images for (a) Wuhan and (b) Daxing (black and white represent the training and remaining prediction samples, respectively).

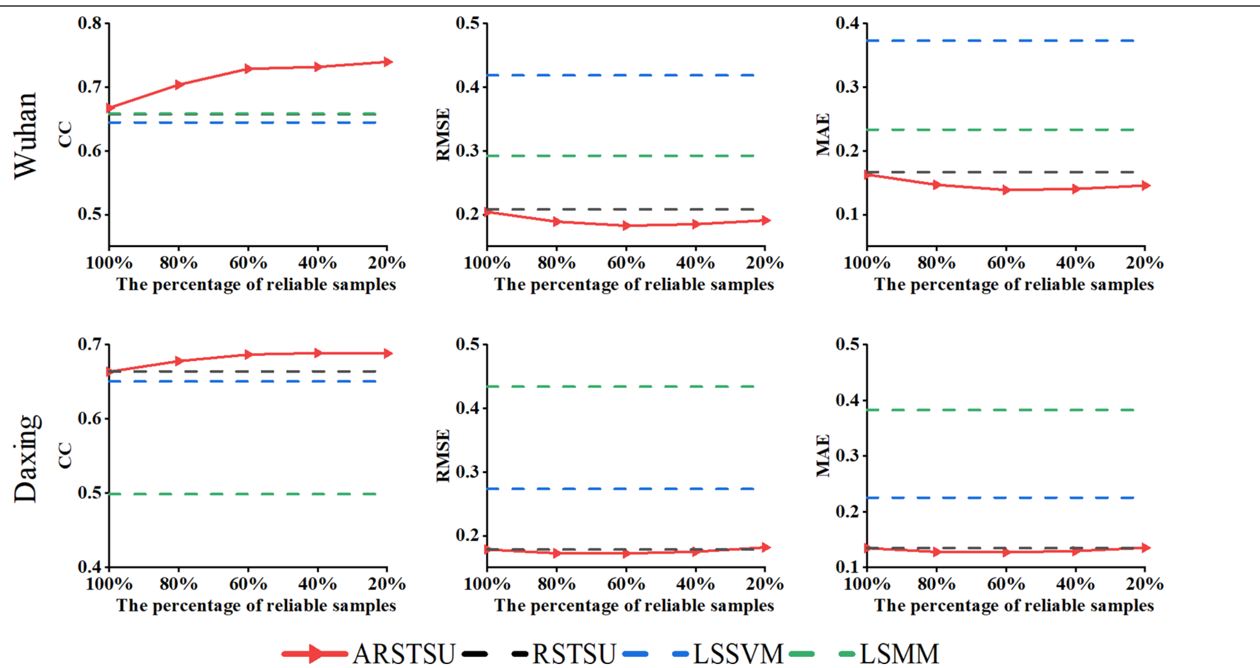


Figure 6. Correlation coefficients (CCs), root-mean-square error (RMSE), and mean absolute error (MAE) values under different percentages of reliable samples for the four spectral-unmixing methods.

whereas the LSSVM result in Daxing is dominated by blue pixels (i.e., almost covered by non-vegetation). Third, the error of RSTSU is smaller than that of LSMM and LSSVM. However, the effect of cloud cover leads to a limited number of training samples extracted, and the result of spectral unmixing is not as accurate as in the ARSTSU model. Based on the selected and augmented samples, a larger number of reliable training samples is produced by ARSTSU, resulting in greater accuracy of spectral unmixing.

Results of the Data Simulated with Real Cloud Masks

In reality, the shape of clouds is more complex. To examine the proposed method in such a case, we simulated the temporally neighboring data with three real cloud masks for the Wuhan data. Figure 10 shows the accuracy evaluation (CCs, RMSE, and MAE) for the four spectral-unmixing methods with different cloud masks. It can be seen that the ARSTSU method has obvious advantages in all cases. For example, in Case 1, when the percentage of reliable samples is 60%, compared with LSMM, LSSVM, and RSTSU, the proposed ARSTSU method decreases the RMSE by 0.09, 0.24, and 0.02, respectively. Moreover, in Case 3, when the percentage of reliable samples is 40%, the corresponding RMSE is 0.10, 0.24, and 0.03 smaller. Thus, ARSTSU is a practical solution for spectral unmixing when the temporally neighboring data are contaminated with different clouds.

Conclusion

Up to now, there have been few studies on spectral unmixing of time-series data based on the goal of dynamic monitoring. The recently proposed STSU model (Q. Wang *et al.* 2021a) enhances the reliability of spectral unmixing by making full use of multi-scale time-series data, which relieves the need for end-member extraction and accounts for intraclass spectral variation. However, STSU is only suitable for processing historical time-series data. Subsequently, the RSTSU model (Q. Wang *et al.* 2021b) was proposed for handling real-time data, which usually uses spatially complete (i.e., without cloud contamination) auxiliary data, leading to a longer temporal distance between effective auxiliary data searched in the temporal domain and real-time data for unmixing. To reduce the effect of long time intervals (e.g., land cover changes) and enhance the reliability of training samples, in this article we proposed the ARSTSU method to select the auxiliary data (almost) temporally closest to the prediction time and to overcome the issue of cloud contamination. Specifically, ARSTSU makes full use of the cloud-free area in the temporal neighbor by selecting and augmenting training samples extracted primarily by change detection based on cloud-free pixels. The augmented samples are then used to train the LSSVM before final proportion prediction. We chose two study regions for experimental validation. The results show that compared with three benchmark methods (LSMM, LSSVM, and RSTSU), the ARSTSU method can produce greater accuracy of spectral unmixing.

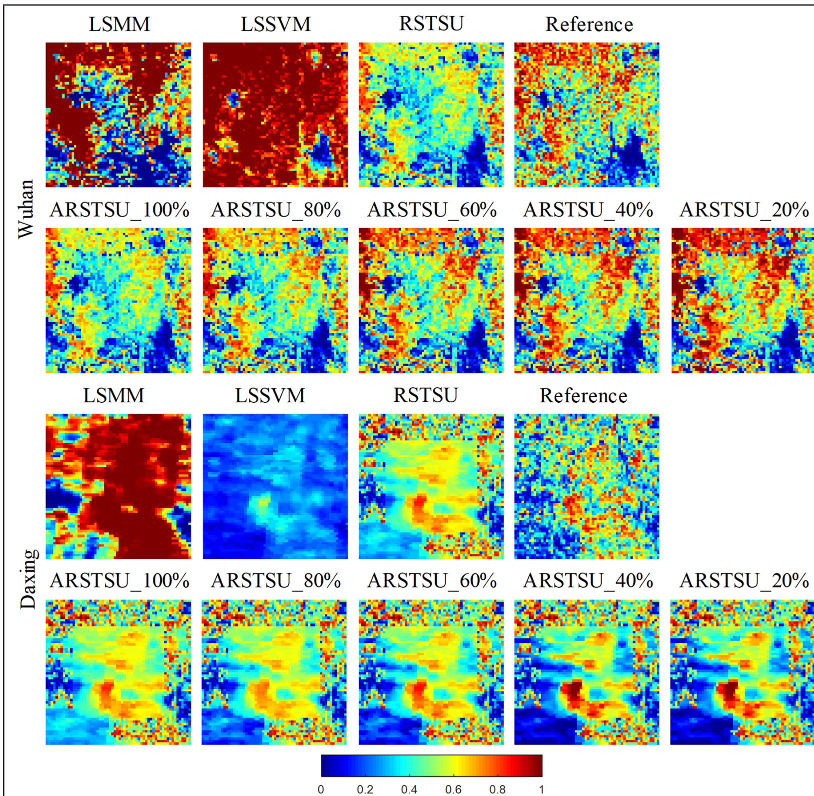


Figure 7. The proportion maps under different percentages of reliable samples for the four spectral-unmixing methods.

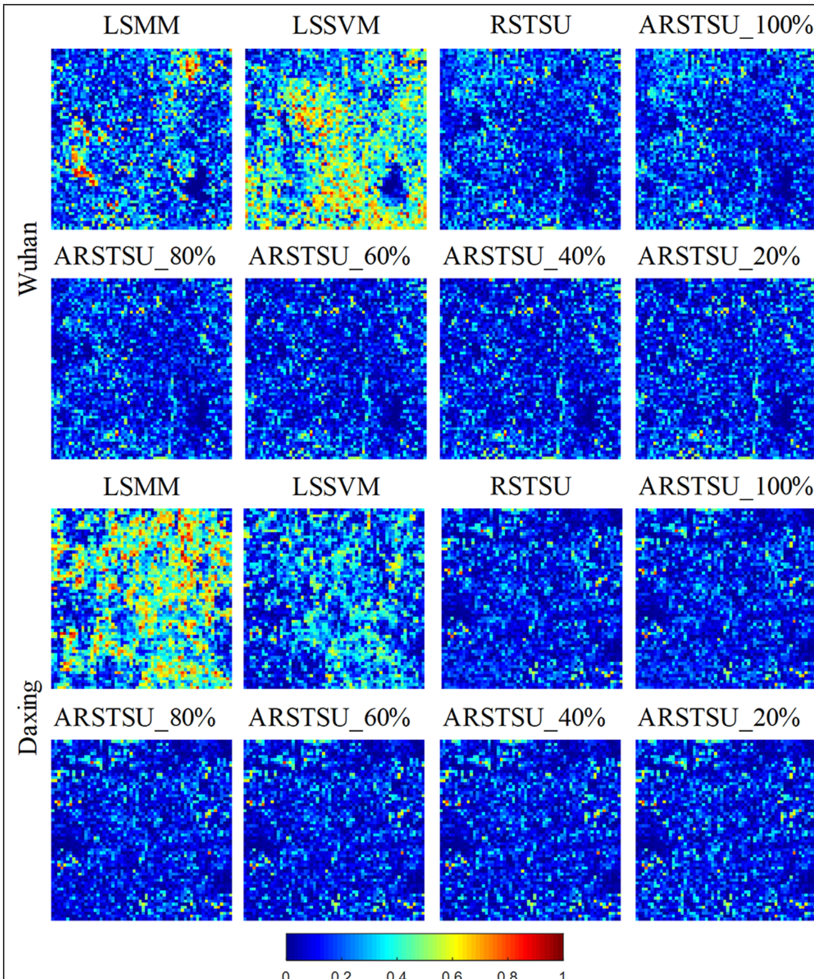


Figure 8. The error maps (in absolute values) under different percentages of reliable samples for the four spectral-unmixing methods.

Acknowledgments
This work was supported by the National Natural Science Foundation of China under grants 42171345 and 41971297, and Tongji University under grant 02502350047.

References

Atkinson, P. M. 2005. Sub-pixel target mapping from soft-classified, remotely sensed imagery. *Photogrammetric Engineering and Remote Sensing* 71(7):839–846.

Bateson, C. A., G. P. Asner and C. A. Wessman. 2000. Endmember bundles: A new approach to incorporating endmember variability into spectral mixture analysis. *IEEE Transactions on Geoscience and Remote Sensing* 38(2):1083–1094.

Bhatt, J. S. and M. V. Joshi. 2020. Deep learning in hyperspectral unmixing: A review. Pages 2189–2192 in *IGARSS 2020–2020 IEEE International Geoscience and Remote Sensing Symposium*, held in Waikoloa, Hawaii, 26 September–2 October 2020. Edited by J. Editor. Location: IEEE.

Chen, B., B. Huang, L. Chen and B. Xu. 2017. Spatially and temporally weighted regression: A novel method to produce continuous cloud-free Landsat imagery. *IEEE Transactions on Geoscience and Remote Sensing* 55(1):27–37.

Deng, C. and Z. Zhu. 2020. Continuous subpixel monitoring of urban impervious surface using Landsat time series. *Remote Sensing of Environment* 238:110929.

Foody, G. M. 2002. The role of soft classification techniques in the refinement of estimates of ground control point location. *Photogrammetric Engineering and Remote Sensing* 68(9):897–903.

Gao, G. and Y. Gu. 2017. Multitemporal Landsat missing data recovery based on tempo-spectral angle model. *IEEE Transactions on Geoscience and Remote Sensing* 55(7):3656–3668.

Goward, S. N., T. V. Loboda, D. L. Williams and C. Huang. 2019. Landsat orbital repeat frequency and cloud contamination: A case study for eastern United States. *Photogrammetric Engineering and Remote Sensing* 85(2):109–118.

Heinz, D. C. and C.-I. Chang. 2001. Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing* 39(3):529–545.

Heylen, R., M. Parente and P. Gader. 2014. A review of nonlinear hyperspectral unmixing methods. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7(6):1844–1868.

Keshava, N. and J. F. Mustard. 2002. Spectral unmixing. *IEEE Signal Processing Magazine* 19(1):44–57.

Li, J., Y. Li, L. He, J. Chen and A. Plaza. 2020. Spatio-temporal fusion for remote sensing data: An overview and new benchmark. *Science China Information Sciences* 63(4):140301.

Lin, A., H. Wu, G. Liang, A. Cardenas-Tristan, X. Wu, C. Zhao and D. Li. 2019. A big data-driven dynamic estimation model of relief supplies demand in urban flood disaster. *International Journal of Disaster Risk Reduction* 49:101682.

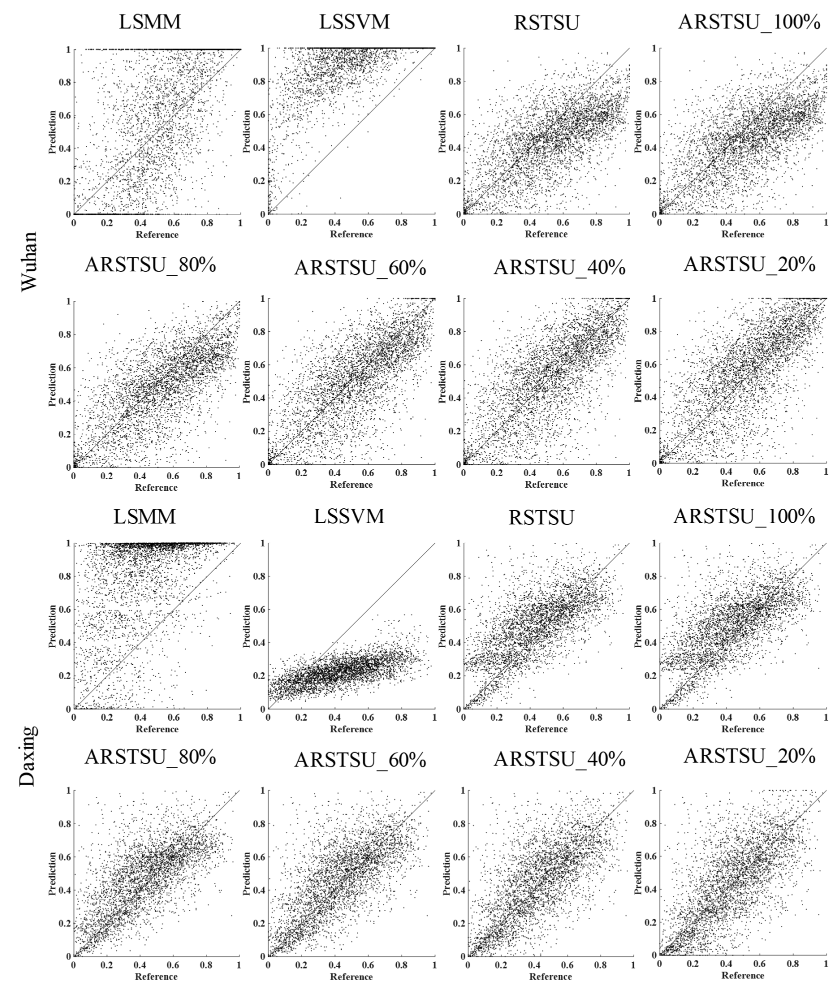


Figure 9. The scatterplots of vegetation proportion (prediction versus reference) for the four spectral-unmixing methods.

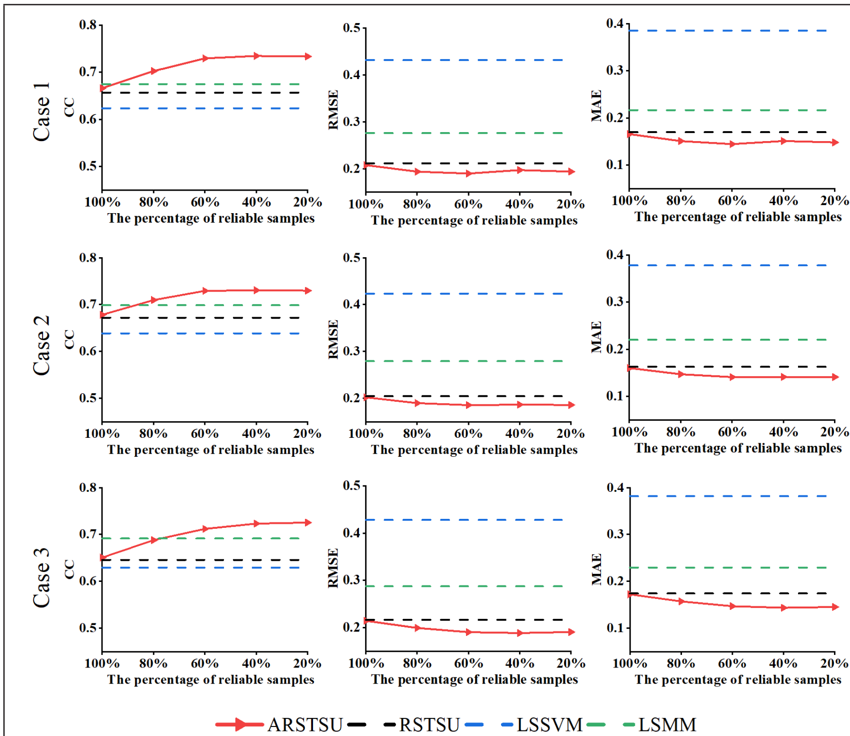


Figure 10. Correlation coefficients (CCs), root-mean-square error (RMSE), and mean absolute error (MAE) values under different percentages of reliable samples for the four spectral-unmixing methods with different cloud masks.

Pal, M. and G. M. Foody. 2010. Feature selection for classification of hyperspectral data by SVM. *IEEE Transactions on Geoscience and Remote Sensing* 48(5):2297–2307.

Quintano, C., A. Fernández-Manso, Y. E. Shimabukuro and G. Pereira. 2012. Spectral unmixing. *International Journal of Remote Sensing* 33(17):5307–5340.

Roberts, D. A., M. Gardner, R. Church, S. Ustin, G. Scheer and R. O. Green. 1998. Mapping chaparral in the Santa Monica Mountains using multiple endmember spectral mixture models. *Remote Sensing of Environment* 65(3):267–279.

Schowengerdt, R. A. 1997. Remote Sensing: Models and Methods for Image Processing. San Diego, Calif.: Academic Press.

Settle, J. J. and N. A. Drake. 1993. Linear mixing and the estimation of ground cover proportions. *International Journal of Remote Sensing* 14(6):1159–1177.

Shen, H., X. Li, Q. Cheng, C. Zeng, G. Yang, H. Li and L. Zhang. 2015. Missing information reconstruction of remote sensing data: A technical review. *IEEE Geoscience and Remote Sensing Magazine* 3(3):61–85.

Shi, C. and L. Wang. 2014. Incorporating spatial information in spectral unmixing: A review. *Remote Sensing of Environment* 149:70–87.

Suykens, J.A.K., J. De Brabanter, L. Lukas and J. Vandewalle. 2002. Weighted least squares support vector machines: Robustness and sparse approximation. *Neurocomputing* 48(1–4):85–105.

Vapnik, V. N. 1995. *The Nature of Statistical Learning Theory*. New York: Springer-Verlag.

Wang, H., J. Li and F. Yang. 2014. Overview of support vector machine analysis and algorithm. *Application Research of Computers* 31(5):1281–1286.

Wang, Q., X. Ding, X. Tong and P. M. Atkinson. 2021a. Spatio-temporal spectral unmixing of time-series images. *Remote Sensing of Environment* 259:112407.

Wang, Q., X. Ding, X. Tong, P. M. Atkinson. 2021b. Real-time spatio-temporal spectral unmixing of remotely sensed images. *IEEE Transactions on Geoscience and Remote Sensing* (submitted).

Wang, Q., K. Peng, Y. Tang, X. Tong and P. M. Atkinson. 2021. Blocks-removed spatial unmixing for downscaling MODIS images. *Remote Sensing of Environment* 256:112325.

Wang, Q., Y. Tang, X. Tong and P. M. Atkinson. 2020. Virtual image pair-based spatio-temporal fusion. *Remote Sensing of Environment* 249:112009.

Wang, Q., L. Wang, C. Wei, Y. Jin, Z. Li, X. Tong and P. M. Atkinson. 2021. Filling gaps in Landsat ETM+ SLC-off images with Sentinel-2 images. *International Journal of Applied Earth Observation and Geoinformation* 101:102365.

Wu, H., A. Lin, K. C. Clarke, W. Shi, A. Cardenas-Tristan and Z. Tu. 2021. A comprehensive quality assessment framework for linear features from Volunteered Geographic Information. *International Journal of Geographical Information Science* 35(9):1826–1847.

Wu, H., Z. Li, K. C. Clarke, W. Shi, L. Fang, A. Lin and J. Zhou. 2019. Examining the sensitivity of spatial scale in cellular automata Markov chain simulation of land use change. *International Journal of Geographical Information Science* 33(5):1040–1061.

Zhang, C., L. Ma, J. Chen, Y. Rao, Y. Zhou and X. Chen. 2019. Assessing the impact of endmember variability on linear Spectral Mixture Analysis (LSMA): A theoretical and simulation analysis. *Remote Sensing of Environment* 235:111471.

Zurita-Milla, R., L. Gomez-Chova, L. Guanter, J.G.P.W. Clevers and G. Camps-Valls. 2011. Multitemporal unmixing of medium-spatial-resolution satellite images: A case study using MERIS images for land-cover mapping. *IEEE Transactions on Geoscience and Remote Sensing* 49(11):4308–4317.

IN-PRESS ARTICLES

Qing Ding, Zhenfeng Shao, Xiao Huang, Orhan Altan, and Yewen Fan. Improving Urban Land Cover Mapping with the Fusion of Optical and SAR Data Based on Feature Selection Strategy.

Clement E. Akumu and Eze O. Amadi. Examining the Integration of Landsat Operational Land Imager (OLI) with Sentinel-1 and Vegetation Indices in Mapping Southern Yellow Pines (Loblolly, Shortleaf and Virginia Pines).

Xinyu Ding and Qunming Wang. Augmented Sample-Based Real-Time Spatiotemporal Spectral Unmixing.

Muhammad Nasar Ahmad, Zhenfeng Shao, and Orhan Altan. Effect of Locust Invasion and Mitigation Using Remote Sensing Techniques: A Case Study of North Sindh Pakistan.

Raechel Portelli, Paul Pope. Remote Sensing and Human Factors Research: A Review.

Weixun Zhou, Wanxuan Geng, Shuanggen Jin. Multi-View Urban Scene Classification With A Complementary Information Learning Model.

Jun Xu, Jiansong Li, Hao Peng, Yanjun He, and Bin Wu. Information Extraction from High-Resolution Remote Sensing Images Based on Multi-Scale Segmentation and Case-Based Reasoning.

Xuehan Wang, Zhenfeng Shao, Xiao Huang, and Deren Li. Spatiotemporal Temperature Fusion Based on a Deep Convolutional Network.

Fan Yang, Zhiwei Fan, Chao Wen, Xiaoshan Wang, Xiaoli Li, Zhiqiang Li, Xintao Wen, and Zhanyu Wei. Three-Dimensional Point Cloud Analysis for Building Seismic Damage Information.

Steven Spiegel, Casey Shanks, and Jorge Chen. Effectiveness of Deep Learning Trained on SynthCity Data for Urban Point-Cloud Classification.

Linze Bai, Qimin Cheng, Yuxuan Shu, and ,Sihang Zhang. Estimating the Aboveground Biomass of Urban Trees by Combining Optical and Lidar Data: A Case Study of Hengqin, Zhuhai, China.

Qimin Cheng, Haiyan Huang, Yin Pan, Neema Nicodemus Lyimo, Hao Peng, Gui Cheng. Cloud Detection in ZY-3 Multi-angle Remote Sensing Images.

Xuzhe Duan, Qingwu Hu, Pengcheng Zhao, and Shaohua Wang. A low-cost and portable indoor 3D mapping approach using biaxial line laser scanners and a one-dimension laser rangefinder integrated with MEMS.

Shaohua Wang, Qingwu Hu, Pengcheng Zhao. A low-cost and portable indoor 3D mapping approach using biaxial line laser scanners and a one-dimension laser rangefinder integrated with MEMS.

Pennan Chinnasamy, Neetu Singh, Shivanand Mallikarjun Nalgire, Meeta Gupta. Potential of open source remote sensing data for improved spatiotemporal monitoring of inland water quality in India: Case study of Gujarat.

Jantien Stoter, Ravi Peters, Balázs Dukai, Stelios Vitalis, Jordi van Liempt. Automated 3D reconstruction of LoD2 and LoD1 models for all 10 million buildings of the Netherlands.

Muhammad Nasar Ahmad, Qimin Cheng, Fang Luo. Dynamic linkage between urbanization, electrical power consumption and suitability analysis using Remote sensing and GIS techniques.

Mahendra Rajitha Udawalpola, Amit Hasan, Anna K. Liljedahl, Aiman Soliman, Jeffrey Terstriep, Chandi Witharna. An Optimal GeoAI Workflow for Pan-Arctic Permafrost Feature Detection from High-Resolution Satellite Imagery.

Zhenfeng Shao, Yan Gu, Xiao Huang, Yuanhao Fu, Jiyuan Gao, Yewen Fan. Assessing the impact of land use changes on net primary productivity in Wuhan, China.

Amit Hasan, Mahendra Udawalpola, Anna Liljedahl, Chandi Witharana. Use of Commercial Satellite Imagery to Monitor Changing Arctic Polygonal Tundra.

Jun Chen, Cunjian Yang, Zhengyang Yu. Research on Machine Intelligent Perception of Urban Geographic Location Based on High Resolution Remote Sensing Images.

Ming Zhang, Lei Zhang, Bowen Wen, Qiongqiong Lan, Qian Wang. An evaluation of pan-sharpening methods for SuperView-1 satellite imagery.

Clement Akumu, Sam Dennis. Urban Land Cover/Use Mapping and Change Detection Analysis Using Multi-Temporal Landsat OLI with Lidar-DEM and Derived TPI.

Guoqing Li, Yingjie Bo, Yelong Zeng, Zhe Liu. Floating solar park impacts urban land surface temperature distribution pattern.

Lihua Wang, lifeng Liang, Benhua Tan, Sicheng Li, Zhiming Kang, Xiujuan Liu. Identifying the driving factors of urban land surface temperature.

Toshihiro Sakamoto, Daisuke Ogawa, Satoko Hiura, Nobusuke Iwasaki. Alternative procedure to improve the positioning accuracy of orthomosaic images acquired with Agisoft Metashape and DJI P4 Multispectral for crop growth observation.

Effect of Locust Invasion and Mitigation Using Remote Sensing Techniques: A Case Study of North Sindh Pakistan

Muhammad Nasar Ahmad, Zhenfeng Shao, and Orhan Altan

Abstract

This study comprises the identification of the locust outbreak that happened in February 2020. It is not possible to conduct ground-based surveys to monitor such huge disasters in a timely and adequate manner. Therefore, we used a combination of automatic and manual remote sensing data processing techniques to find out the aftereffects of locust attack effectively. We processed MODIS-normalized difference vegetation index (NDVI) manually on ENVI and Landsat 8 NDVI using the Google Earth Engine (GEE) cloud computing platform. We found from the results that, (a) NDVI computation on GEE is more effective, prompt, and reliable compared with the results of manual NDVI computations; (b) there is a high effect of locust disasters in the northern part of Sindh, Thul, Ghari Khairo, Garhi Yaseen, Jacobabad, and Ubauro, which are more vulnerable; and (c) NDVI value suddenly decreased to 0.68 from 0.92 in 2020 using Landsat NDVI and from 0.81 to 0.65 using MODIS satellite imagery. Results clearly indicate an abrupt decrease in vegetation in 2020 due to a locust disaster. That is a big threat to crop yield and food production because it provides a major portion of food chain and gross domestic product for Sindh, Pakistan.

Introduction

Locusts are a species of insects that are and have been threatening food security in the history of humankind. Many international organizations are working on the prevention of locust plagues because it directly damages massive crop and cultivated areas. In a report of the United Nations Food and Agriculture Organization (FAO), locusts travel in clusters (Ma *et al.* 2005) with a count of approximately more than 80 million per square kilometer. It is an overall estimate that 10% of the livelihood (Latchininsky 2013) of the world population is affected by locust disaster each year. FAO also initiated the Desert Locust Plague Prevention Programme, which supports information on a global scale and gives early warning systems on agriculture and food security (Hielkema *et al.* 1986).

To prevent the locust invasion, early identification, prompt control, and continuous information of locust population (Michael *et al.* 2017) is required along with the condition of vegetation and rainfall data. Sufficient precipitation gives the necessary soil moisture for supporting the advancement of dense vegetation (Hielkema and Snijders 1994), which is used by the locusts as a source of food and shelter.

Remote sensing satellite products are permitting researchers to monitor locust disasters using both moderate- and high-resolution imagery. Remote sensing satellite products are suitable to identify locust populations on a large topographical scale for estimating areas invaded due to locust disaster (Ji *et al.* 2004). The use of satellite products for monitoring locust habitat is very advantageous (Bryceson and Wright 1986; Justice *et al.* 1985; Tappan *et al.* 1991). It involves determining reflected interaction between spectral radiance and green-leaf vegetation cover (Tuckler *et al.* 1985) that can be extracted from satellite data. Dynamic properties of satellite imagery have proved that it has the potential to detect the effects of locusts on vegetation and for determining the crop losses (Dottavio and Williams 1983; Rencz and Nemeth 1985; Wewetzer *et al.* 1993).

The Advanced Very High-Resolution Radiometer (AVHRR) can provide exceptional results compared with other satellite products. AVHHR sensors are based on environmental polar-orbiting that is provided by the National Oceanic and Atmospheric Administration (NOAA) and has a significant capability to identify such ecological phenomenon on a regional scale (Hielkema and Snijders 1994). However, there is a limitation on data availability because very high-resolution and high-frequency data is required to monitor locust movement, which can be achieved by installing more satellites to enhance the capability of satellite imagery.

In moderate-resolution satellites and open-source data, normalized difference vegetation index (NDVI) has more potential for early detection and monitoring of epidemic regions (Cherlet and Hielkema 1989; Ma *et al.* 2005). In a report, the Australian Plague Locust Commission discussed the strength of NDVI products because it principally focuses on red and near-infrared spectral bands, which is very speedy and efficient (Kriegler *et al.* 1969) for the detection of vegetation change. MODIS is providing NDVI products with a high temporal resolution that covers the globe and monitor Earth's vegetation activities (Ji *et al.* 2004; Huete *et al.* 1999) in 1–2 days. It is available in a moderate spatial resolution from 250 meters to 1 kilometer. That supports research applications primarily related to vegetation loss and change detection. For example, dense vegetation has a higher reflectance value in the near-infrared region as compared with the green area. Additionally, more applications are necessary to control the locust attack on a large-scale (Latchininsky and Sivanpillai 2010; Beck *et al.* 2006) because it also affects our environment and can be more harmful in future.

Muhammad Nasar Ahmad and Zhenfeng Shao are with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China (mnasarahmad@whu.edu.cn).

Orhan Altan is with the Department of Geomatics, Istanbul Technical University, Istanbul, 36626, Turkey.

Contributed by Hongyan Zhang, May 5, 2021 (sent for review August 2, 2021; reviewed by Timothy A. Warner, Nana Yaw Danquah Twumasi, Neema S. S. Sumari).

Photogrammetric Engineering & Remote Sensing
Vol. 88, No. 1, January 2022, pp. 47–53.
0099-1112/22/47–53

doi: 10.14358/PERS.21-00025R2

In this case study, the authors are focused on (a) combining automatic and manual remote sensing techniques to detect the effect of locust attacks, (b) processing MODIS-NDVI on ENVI and *Landsat 8* NDVI on the Google Earth Engine (GEE) platform, (c) performing analysis for vulnerable cities within the study area, and (d) suggesting improved ways to mitigate future locust disasters. For this purpose, both classification techniques have been adopted, including manual and automatic image classification. The authors used *Landsat 8* OLI satellite imagery on the GEE platform to identify change due to locust invasion in north Sindh. Basically, GEE is a cloud-based geospatial analysis platform that makes it really easy for researchers to access huge amounts of data sets with the help of high performance computational capabilities. This allows the authors to study various global issues like land use, land cover change detection, deforestation, weather monitoring, and environmental change. It is a highly dedicated platform that makes trouble-free processing of remote sensing data sets with the help of a programming language. The user, after creating a workable algorithm on the platform, can integrate with applications, websites, and other platforms as well. Most of the data set is available including Landsat Archive, MODIS, and Sentinel (Gorelick *et al.* 2017) in the Google Earth Engine data catalog.

The programming compatibility with GEE allows for creating any algorithm, indexes, and programs. It can support multiple analysis at a time to process, which greatly increases the overall work speed (Kumar and Mutanga 2018). Therefore, GEE is a very effective platform that allows the authors to use massive remote sensing data sets through an online cloud-based approach, performing intensive analysis and process on data sets (Shelestov *et al.* 2017).

Pakistan has less desert locust swarm history compared with Africa and many other countries in the world, Desert locusts have more capacity to damage crops quickly compared with other species of locusts. In the South-Asia region, due to heavy precipitation, availability of wet soil in deserts supports locusts' survival, allowing them to lay eggs and to reproduce their groups. Historically, the earliest epidemic of locust observed in Karachi was in 1961. However, in 2020, Pakistan faced the worst desert swarms to date, with invasions in Punjab, Sindh, and Baluchistan. These swarms are also a short distance from the Xinjiang border of China.

Desert Locusts were also discovered in 2019 but were under control, and there was no significant damage. However, in 2020, a devastating outbreak was observed because of heavy rainfall in the monsoon season. That provided an ideal situation for the production of locust swarms in agricultural land because of dense vegetation. Locusts entered first in the deserts of Punjab province, which is located in Yazman Mandi, and moved towards North Sindh and Baluchistan. According to statistics of the Ministry of National Food Security and Research Pakistan, about 80 000 hectares of crops were demolished due to locust attack in Sindh. Pakistan has two seasons that support locusts for breeding as compared with other countries.

Desert locusts can fly very fast over long distances and can fly over a 150 km area in a day. According to an estimate, about 40 million locusts are present in a locust swarm. FAO is currently working on the mapping of the migration of locusts in the Arabian Peninsula, East Africa, and Southwest Asia. In February 2020, locusts moved in a southwesterly direction from the south desert area of Pakistan. However, in March 2020, locust groups theatrically increased in the northern part of Sindh Province and in April 2020, the both the desert locust group settlements and bands grew. (Yamano *et al.* 2020).

The government of Pakistan declared an emergency in February 2020 to control the locust swarms. Still, due to favorable weather conditions, locusts bred very fast and attacked a major part of the crop in North Sindh. An estimated loss of $6 million in wheat was recorded in Sindh alone due to locust outbreak and a threat to cotton, rice, millet, corn, and sorghum was ongoing. There is a high risk to food security and production because the value is 18.5% of the gross domestic product of Pakistan.

This research contributes to the monitoring of locust invasion in selected areas in North Sindh, Pakistan. No articles have been published on locust assessment using MODIS NDVI and Landsat NDVI using the GEE platform in this study area.

Study Area and Data

Study Area

Sindh is the more important province of Pakistan in terms of agricultural land and provides a major portion of agriculture food production. It is a densely populated province, but in this research, only the northern part of Sindh is focused on. The geographical coordinates are 27°57'21.0"N latitude and 68°37'19.1"E longitude. Figure 1 shows the extent of the study area.

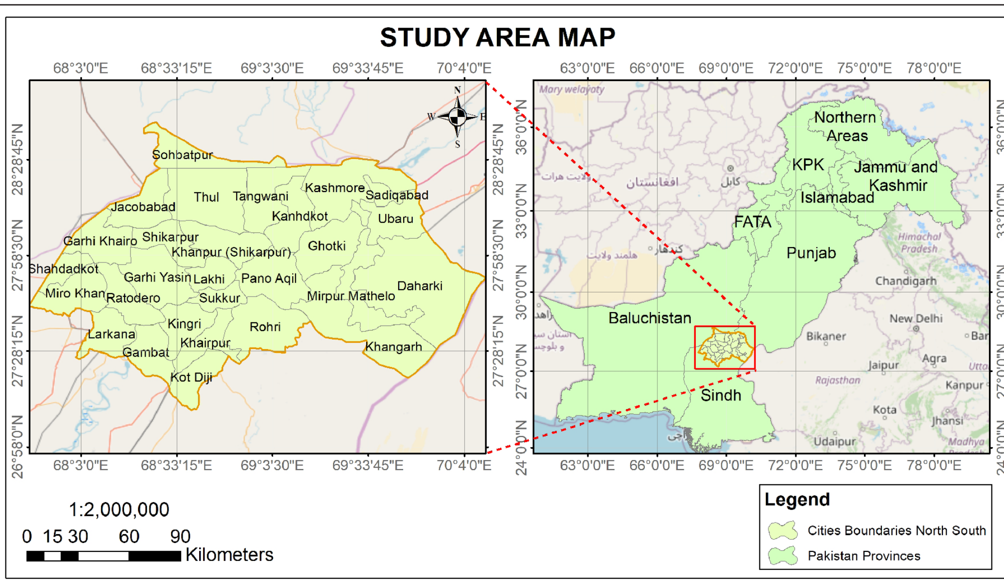


Figure 1. Location of the study areas on the map.

Table 1. Description of satellite imagery used.

Satellite or Instrument Name	Sensor	Bands	Acquisition Date	Spatial Resolution	Source
MODIS	MOD13Q1	Bands 1 (red), 2 (near-infrared), 3 (blue), and 7 (mid-infrared)	February 2019, 2020	250 m	https://lpdaac.usgs.gov/products/mod13q1v006/
Landsat 8	OLI-TRIS	Band (1-8), near-infrared (9), thermal bands	February 2019, 2020	30 M	https://earthengine.google.com/

Data Set Overview

MODIS NDVI

In this research, the MOD13Q1 product is used that includes two vegetation layers: enhanced vegetation index and normalized difference vegetation index (Gill *et al.* 2009). Both were acquired at a 250-meter resolution over a 16-day time interval. According to Mkhabela *et al.* (2011), AVHHR offers high-resolution, six band multispectral data that is used to monitor crop conditions and computation of NDVI indices. Portability of intermediate data products and results can also be an additional advantage in the remote sensing scientific community (Lunetta *et al.* 2006). There is no previous research carried out in the Sindh Provinces using the MOD13Q1 product for locust monitoring. AVHHR-NDVI indicates the pixel value with very high accuracy; the algorithm is based on the highest NDVI value, low clouds, and view angle.

LANDSAT 8

Landsat imagery was acquired on the GEE platform under the National Aeronautics and Space Administration (NASA)/USGS program, which provides the longest and most continuous space-based satellite imagery worldwide. GEE provided Landsat 5 imagery from 1985 to 2011 (Huang *et al.* 2017). This study was carried out using *Landsat 8* OLI imagery for the years 2019 and 2020. The selection of satellite imagery needs to be of a similar date or at least the same month for classification. Table 1 shows the acquisition time and respective platforms for both MODIS and Landsat used in this study.

Methodology

Satellite data can be combined to provide temporally dense observations, or more information for environmental applications, such as for the monitoring of land use/land cover management and their settings (Shao *et al.* 2019). This study hypothesizes that a methodology can be developed to assess vegetation loss in the north Sindh due to locust plague by using multi-resolution satellite data and multiple platforms. The data flow diagram in Figure 2 shows the overall procedure of the analysis.

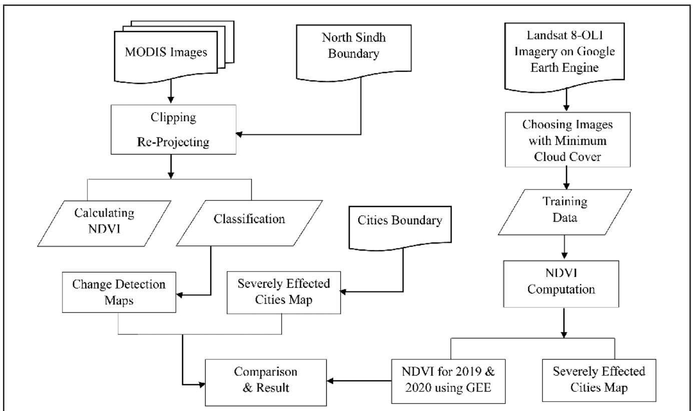


Figure 2. Data flow diagram.

The total area of vegetation affected due to locust attack is extracted from the classified image to analyze, quantify, and to visualize the obtained results for further computations. For instance, the total vegetation area for both years is extracted using Equation 1. NP designates the number of pixels for vegetation class and PA the total pixel area, which is always 0.06 km^2 for MODIS MOD13Q1 because it has a resolution of 250 meters (Zhang *et al.* 2017).

$$\text{Area} = \text{NP} \times \text{PA} \times 10^{-6} \frac{\text{km}^2}{\text{m}^2} \tag{1}$$

NDVI can be calculated by dividing the near-infrared (NIR) band and red (R) band. If there is stress on vegetation, it changes reflectance values in the reverse direction. The NDVI focus on the density and greenness of vegetation (Tarpley *et al.* 1984) that is calculated as shown in Equation 2.

$$\text{NDVI} = \frac{\text{NIR} - \text{R}}{\text{NIR} + \text{R}} \tag{2}$$

where NIR and R represent the reflectance of the near-infrared and red band, respectively (Mkhabela *et al.* 2011).

Image Classification Method and Details

MODIS-MOD13Q1 satellite imagery is acquired and has been classified using a support vector machine (SVM) for 2019 and 2020, respectively. The SVM is a supportive classifier for spectral detection and a highly effective supervised classifier (Li *et al.* 2012). Satellite images are further processed to analyze the change in NDVI values for the locust year that is explained in results below.

Google Earth Engine Cloud Computing Platform

In terms of GEE, the methodology of this research work includes a selection of satellite data, feature classes, and performing supervised classification. All these steps are performed on GEE, cloud-based processing using remote sensing imagery of study area. GEE code and script were generated by the authors to carry out the classification and analysis. Firstly, satellite imagery of the study area was loaded on the GEE platform and some sample points were constructed manually to make classification results more reliable and accurate.

Landsat 8 images were used for performing classification and NDVI time series for 2019 and 2020. Google Earth Engine provides an open access to the Landsat archive. It includes satellite imagery from *Landsat 1–Landsat 8*. For this research, *Landsat 8* imagery was only used for the above mentioned years. As can be seen in Figure 3, the primary NDVI output is for North Sindh, Pakistan.

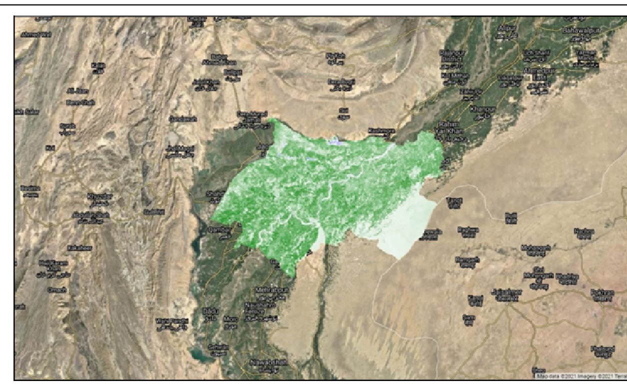


Figure 3. Preview of *Landsat 8* classified image on GEE.

Classification Scheme

After the acquisition of raw satellite imagery, proper corrections have been made to make them useful for further analysis. The imagery is then used for analysis-like classification, which is a scheme for extracting information about land cover and land use attributes. For selection of the method for classification, it totally depends upon the objective and the available resources. The human eye has a very sensitive approach to color difference and object identification. However, it is a hectic and time-consuming task that is greatly dependent on the experience of the observer. In addition to this, manual interpretation is not possible when a large area and multiple spectral bands are involved. In order to handle this problem, computer-aided classification is used; it divides the pixels into specific classes and can be used to perform classification of multispectral or even hyperspectral imagery (Borra *et al.* 2019).

Algorithm

GEE provides an interactive platform with high-end capabilities and high calculation rate using many types of classification algorithms. The most commonly used classification algorithms are Maximum Likelihood Classifier, SVM, and Random Forest (RF). RF has been used in this case study because it is a well-known algorithm (Kumar *et al.* 2019). For the classification, a pixel-based RF machine learning algorithm is used because it is extremely useful in remote sensing data classification on GEE. In addition to this, RF has the ability to handle data with high variation and provide results with high accuracy compared with other approaches. RF constructs random decision trees and collects them to classify the data from results of all decision trees. It is more powerful than a single decision tree and is much easier to use compared with other classifiers (Oliphant *et al.* 2019).

City-Wide Analysis

It is essential to analyze profoundly affected areas further; therefore, a city wide comparison has been done using MODIS and Landsat data. For this purpose, four cities, including Thul, Garhi Khairo, Garhi Yasin, and Jacobabad from North Sindh are further magnified using NDVI processed on ENVI and GEE. It can be observed from Figure 6 and Figure 9 that most of the disaster happened in the northern part of the selected study areas.

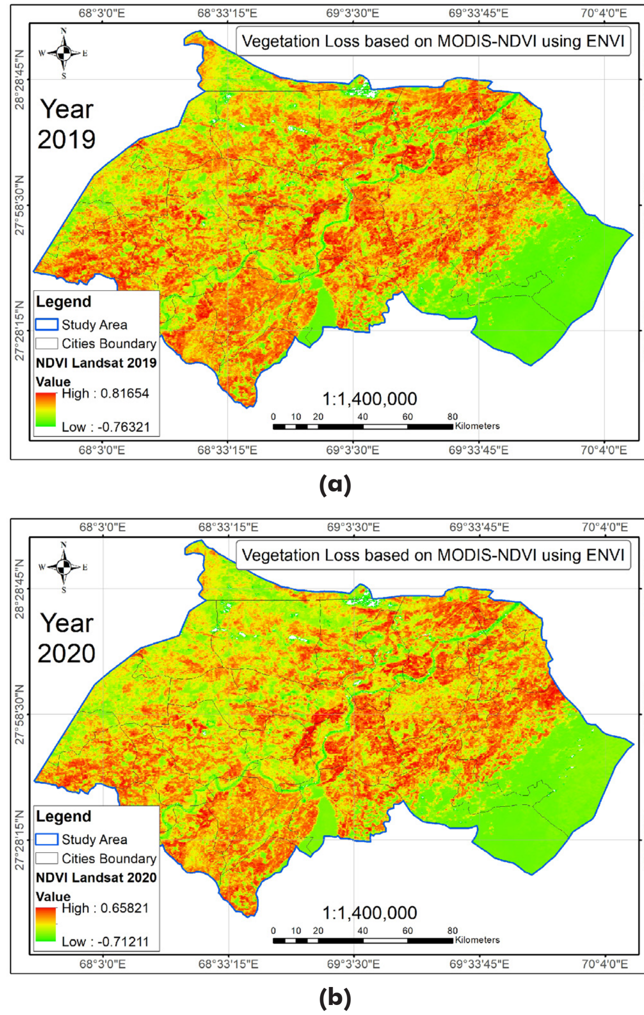


Figure 4. Change in vegetation area, North Sindh using MODIS NDVI. (a) 2019. (b) 2020.

Results

Results are showing most of the vegetation loss due to locust outbreaks observed in the northern parts of the study area. The results are further explained below for each data set.

MODIS-NDVI

A significant plague within one year in the vegetation area can be seen in Figure 4a and 4b, which illustrates the MODIS NDVI and classification results using the SVM classifier for selected years. Overall, change detection can be seen clearly using NDVI results, showing that there is a major decrease in vegetation for 2020. For more precise identification of locust effects on vegetation and difference, some extremely affected cities from the study area are further magnified.

Table 2 shows the overall accuracy and kappa coefficient for North Sindh. A confusion matrix cross validates classifications and determines whether a classifier has been appropriately identified.

Table 2. Overall accuracy and kappa coefficient using MODIS-NDVI.

Accuracy Assessment	North Sindh	
Year	2019	2020
Overall accuracy (%)	89.19%	90.48%
Kappa coefficient	82.85	85.2

It is available in all image classification software to determine classification accuracy. The backend algorithm gives an estimation of the values assigned to a class and if they were accurately classified by using a comparison signature that was made on the same image by the user. As a result, Figure 5 shows how much vegetation is affected due to locust disaster in selected cities including Thul, Garhi Khairo, Garhi Yasin, and Jacobabad. All of these municipalities are located in the northern part of Sindh Province.

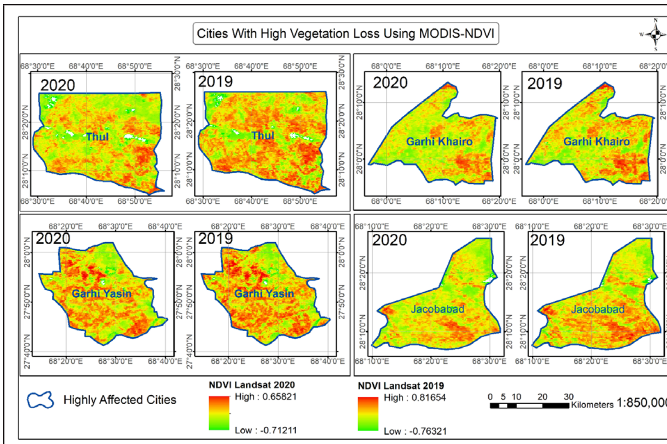


Figure 5. Areas with high vegetation loss using MODIS NDVI.

The bar chart in Figure 6 shows the overall change in major cities for Sindh province, which includes 28 municipalities of North Sindh. For each city, there is massive difference in vegetation change detection in 2020, which is the locust year as compared with vegetation cover in 2019. The bar chart for 2019 and 2020 vegetation is shown on the same bar graph for better representation of vegetation loss.

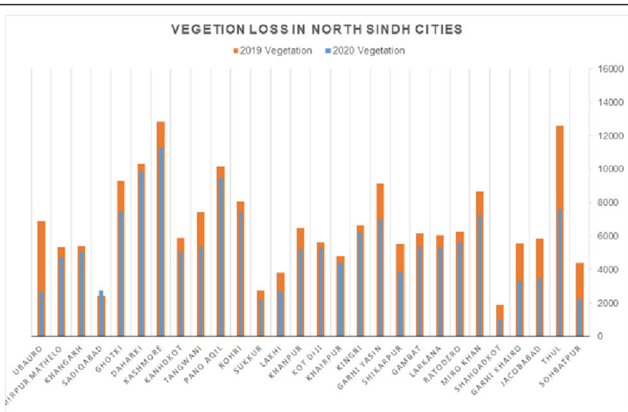


Figure 6. Municipalities are more valuable in North Sindh.

These results are obtained by classification for each city manually for both years. After that, classification results were subtracted using the ArcMap raster analyst tool. Then the authors found that Thul, Ghari Khairo, Garhi Yaseen, Jacobabad, and Ubauro were more effected cities due to locust invasion. It can be seen from Figures 5, 6, and 8 that using MODIS-NDVI and Landsat-NDVI classification results, similarity in results were found.

Landsat NDVI Using Google Earth Engine

From the results, similar patterns were identified using automatic NDVI computation using a programming algorithm on Google Earth Engine. Figure 7a and 7b show the classification results for 2019 and 2020 using Landsat NDVI-GEE.

GEE provided more prompt and consistent results compared with MODIS manual NDVI computations because of spatial resolution, processing time, and machine learning algorithms.

Furthermore, the same cities are taken into consideration to compare with the results of MODIS NDVI as discussed above. There were almost 28 municipalities that were covered in the study area, but the authors only acquired four cities that are

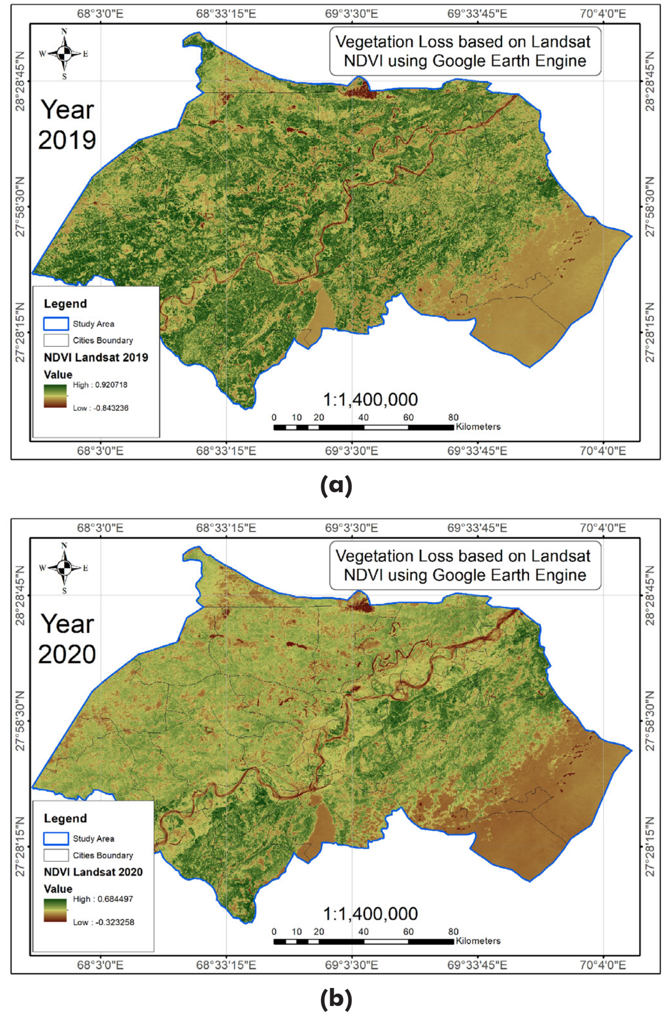


Figure 7. Vegetation change based on Landsat NDVI using GEE. (a) 2019. (b) 2020.

further mapped for Landsat NDVI computation on GEE. The results are shown in Figure 8. It can be seen that NDVI value has an abrupt change from 0.92 to 0.68.

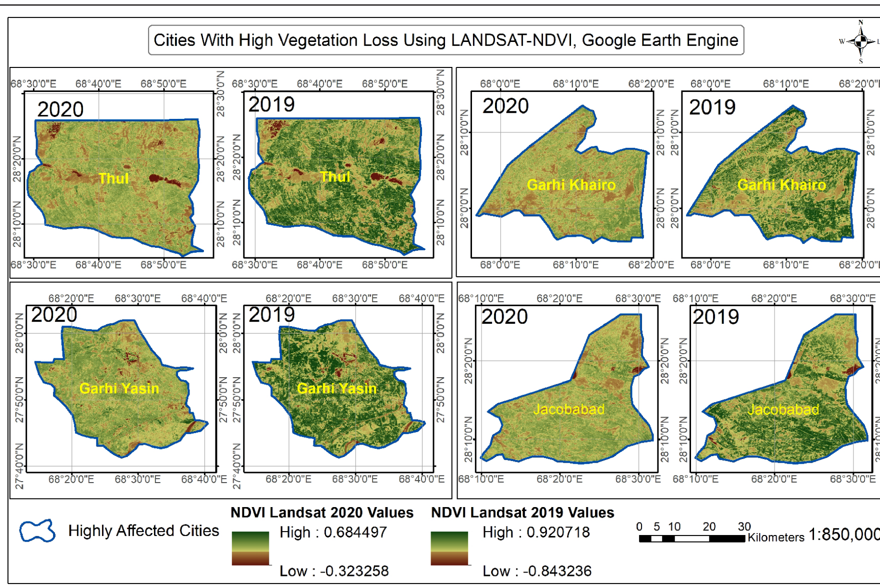


Figure 8. Most vulnerable cities in North Sindh using MODIS NDVI.

Discussion

It is very important to map locust invasion because Sindh provides the major portion of agricultural commodities, including crops and vegetables such as wheat, rice, sugarcane, maize, sorghum, millet, and chickpeas. Additionally, it grows vegetables and fruit such as onion, potato, bottle gourd, cauliflower, eggplant, chilies, ladyfingers, and green pears. The sowing timespan for most of these yields are February to March. Analysis was performed to identify vulnerable cities within the study area and to make food sustainability-driven policies. In 2020, locust outbreak also occurred in Somalia, Yemen, Iran, Egypt, India, Oman, and Tanzania along with Pakistan. So, this case study can assist researchers from these regions as well. This study has shown that satellite-NDVI products can be used effectively to predict vegetation yields. FAO also believes that meteorological satellites used to monitor the environment can detect locust swarms on a regional scale.

However, it is difficult for both national and international locust organizations to interpret the hundreds of images produced every day. There is limitation in terms of (a) high-resolution satellite data availability, (b) advance technology in Ultra Low Volume (ULV) sprayers and fogger machines, and (c) automatic systems based on artificial intelligence and machine learning.

Thus, it is recommended that there should be more high-resolution NDVI products freely available during any locust epidemic. An automatic system must be designed to map locust invasion using programming-based algorithms in the future. The government of Pakistan should acquire more ULV vehicle-mounted sprayers and fogger machines for aerial spraying during a locust outbreak. It is very important to control the desert locust invasion, as it has already affected livelihood, nutrition, and food security to huge populations of Pakistan. In future, there should be improved grassroots assistance for emergency response and building a strong line of defense for locust disaster, prevention, mitigation, and relief.

Conclusions

It is concluded that the locust plague has a more significant impact on agricultural land in Sindh, as shown in the results. In Sindh Pakistan, 80% of the farmers used to cultivate their crops in March and a locust invasion happened in the first quarter of 2020. This paper concludes that NDVI computation on GEE is more effective compared with manual computations. Thul, Ghari Khairo, Garhi Yaseen, Jacobabad, and Ubauro are found to be more vulnerable areas because of locust disaster. NDVI values fall to 0.68 from 0.92 in 2020 using Landsat NDVI and, vice versa 0.81 to 0.65 using MODIS data. Moreover, in the absence of high-resolution satellite images and NDVI products to detect insect invasion in real-time, researchers can find ways to make use of the freely available resources to obtain as close as possible results for analysis and future policy drafting.

Acknowledgments

This work is supported by the National Key R&D plan on strategic international scientific and technological innovation cooperation special project under Grant 2016YFE0202300, the National Natural Science Foundation of China under Grants 41771452, 41771454, and 41890820, and the Natural Science Fund of Hubei Province in China under Grant 2018CFA007. The authors would like to thank the satellite imagery providers NOAA/NGDC and NASA, which are freely available for researchers. The authors would like to thank the anonymous reviewers for their constructive comments and suggestions for improving the manuscript.

References

Beck, P. S., C. Atzberger, K. A. Høgda, B. Johansen and A. K. Skidmore. 2006. Improved monitoring of vegetation dynamics at very high latitudes: A new method using MODIS NDVI. *Remote sensing of Environment* 100(3):321–334.

Borra, S., R. Thanki and N. Dey. 2019. *Satellite Image Analysis: Clustering and Classification*. Singapore: Springer.

Bryceson, K. P. and D. E. Wright. 1986. An analysis of the 1984 locust plague in Australia using multitemporal Landsat multispectral data and a simulation model of locust development. *Agriculture, Ecosystems and Environment* 16(2): 87–102.

Cherlet, M. R. and J. U. Hielkema. 1989. Overview of operational remote sensing techniques and planned developments for supporting desert locust monitoring and forecasting. In *Research Meeting on Desert Locust Forecasting and Populations Dynamics*. Rome, Italy: United Nations Food and Agriculture Organization (FAO).

Dottavio, C. L. and D. L. Williams. 1983. Satellite technology: An improved means for monitoring forest insect defoliation. *Journal of Forestry* 81(1): 30–34.

Gill, T. K., S. R. Phinn, J. D. Armston and B. A. Pailthorpe. 2009. Estimating tree-cover change in Australia: Challenges of using the MODIS vegetation index product. *International Journal of Remote Sensing* 30(6): 1547–1565.

Gorelick, N., M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau and R. Moore. 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment* 202: 18–27.

Hielkema, J. U., J. Roffey and C. J. Tucker. 1986. Assessment of ecological conditions associated with the 1980/81 desert locust plague upsurge in West Africa using environmental satellite data. *International Journal of Remote Sensing* 7(11): 1609–1622.

Hielkema, J. U. and F. L. Snijders. 1994. Operational use of environmental satellite remote sensing and satellite communications technology for global food security and locust control by FAO: The ARTEMIS and DIANA systems. *Acta Astronautica* 32(9): 603–616.

Huang, H., Y. Chen, N. Clinton, J. Wang, X. Wang, C. Liu, P. Gong, J. Yang, Y. Bai, Y. Zheng and Z. Zhu. 2017. Mapping major land cover dynamics in Beijing using all Landsat images in Google Earth Engine. *Remote Sensing of Environment* 202: 166–176.

Huete, A., C. Justice and W. Van Leeuwen. 1999. MODIS vegetation index (MOD13). *Algorithm Theoretical Basis Document* 3(213): 295–309.

Ji, R., B. Y. Xie, D. M. Li, Z. Li and X. Zhang. 2004. Use of MODIS data to monitor the oriental migratory locust plague. *Agriculture, Ecosystems and Environment* 104(3): 615–620.

Justice, C. O., J.R.G. Townshend, B.N. Holben and E.C. Tucker. 1985. Analysis of the phenology of global vegetation using meteorological satellite data. *International Journal of Remote Sensing* 6(8): 1271–1318.

Kriegler, F. J., W. A. Malila, R. F. Nalepka and W. Richardson. 1969. Preprocessing transformations and their effects on multispectral recognition. *Remote Sensing of Environment* VI: 97.

Kumar, L. and O. Mutanga, eds. 2019. *Google Earth Engine Applications*. Basel, Switzerland: MDPI Books, 420p.

Kumar, L. and O. Mutanga. 2018. Google Earth Engine applications since inception: Usage, trends, and potential. *Remote Sensing* 10(10): 1509.

Latchininsky, A. V. 2013. Locusts and remote sensing: A review. *Journal of Applied Remote Sensing* 7(1): 075099.

Latchininsky, A. V. and R. Sivanpillai. 2010. Locust habitat monitoring and risk assessment using remote sensing and GIS technologies. In *Integrated Management of Arthropod Pests and Insect Borne Diseases*, 163–188. Dordrecht, The Netherlands: Springer.

Li, Z., P. Liu, W. Wang and C. Xu. 2012. Using support vector machine models for crash injury severity analysis. *Accident Analysis and Prevention* 45: 478–486.

Lunetta, R. S., J. F. Knight, J. Ediriwickrema, J. G. Lyon and L. D. Worthy. 2006. Land-cover change detection using multi-temporal MODIS NDVI data. *Remote Sensing of Environment* 105(2): 142–154.

Ma, J., X. Han, Hasibagan, C. Wang, Y. Zhang, J. Tang, Z. Xie and T. Deveson. 2005. Monitoring East Asian migratory locust plagues using remote sensing data and field investigations. *International Journal of Remote Sensing* 26(3): 629–634.

Michael, C., M. Pierre, B. Etienne and D. Pierre. 2017. Spot Vegetation Contribution to Desert Locust Habitat Monitoring. <https://www.spot-vegetation.com/pages/vgtprep/vgt2000/cherlet.pdf>

Mkhabela, M. S., P. Bullock, S. Raj, S. Wang and Y. Yang. 2011. Crop yield forecasting on the Canadian Prairies using MODIS NDVI data. *Agricultural and Forest Meteorology* 151(3): 385–393.

Oliphant, A. J., P. S. Thenkabail, P. Teluguntla, J. Xiong, M. K. Gumma, R. G. Congalton and K. Yadav. 2019. Mapping cropland extent of Southeast and Northeast Asia using multi-year time-series Landsat 30-m data using a random forest classifier on the Google Earth Engine Cloud. *International Journal of Applied Earth Observation and Geoinformation* 81: 110–124.

Rencz, A. N. and J. Nemeth, 1985. Detection of mountain pine beetle infestation using Landsat MSS and simulated thematic mapper data. *Canadian Journal of Remote Sensing* 11(1): 50–58.

Shao, Z., J. Cai, P. Fu, L. Hu and T. Liu. 2019. Deep learning-based fusion of *Landsat-8* and *Sentinel-2* images for a harmonized surface reflectance product. *Remote Sensing of Environment* 235: 111425.

Shelestov, A., M. Lavreniuk, N. Kussul, A. Novikov and S. Skakun. 2017. Exploring Google Earth Engine platform for big data processing: Classification of multi-temporal satellite imagery for crop mapping. *Frontiers in Earth Science* 5: 17.

Tappan, G. G., D. G. Moore and W. I. Knausenberger. 1991. Monitoring grasshopper and locust habitats in Sahelian Africa using GIS and remote sensing technology. *International Journal of Geographical Information System* 5(1): 123–135.

Tarpley, J. D., S. R. Schneider and R. L. Money. 1984. Global vegetation indices from the NOAA-7 meteorological satellite. *Journal of Climate and Applied Meteorology* 23(3): 491–494.

Wewetzer, A., S. Krall and F. A. Schulz. 1993. *Methods for the Assessment of Crop Losses Due to Grasshoppers and Locusts*. Eschborn, Germany: Deutsche Gesellschaft fur Technische Zusammenarbeit (GTZ) GmbH.

Yamano, T., N. Sato and B. W. Arif. 2020. *Impact of COVID-19 and Locust Swarms on Farm Households in Sindh, Pakistan: Analysis of Data from a Cross-Sectional Survey*. ADB Briefs, 12p. <http://dx.doi.org/10.22617/BRF200280-2>

Zhang, L., Q. Weng and Z. Shao. 2017. An evaluation of monthly impervious surface dynamics by fusing Landsat and MODIS time series in the Pearl River Delta, China, from 2000 to 2015. *Remote Sensing of Environment* 201: 99–114.

Call for Submissions

AI-Based Environmental Monitoring with UAV Systems

Photogrammetric Engineering and Remote Sensing (*PE&RS*) is seeking submissions for a special issue on AI-Based Environmental Monitoring with UAV Systems.

Global warming and climate change have become the most important factor threatening the world. Climate change results in dramatical environmental hazards and threatens the planet and human life. A wide variety of policies have been proposed to decrease the effects of global warming and climate change. The most important one is the Paris Agreement which aims to limit global warming to well below two degrees Celcius. Many countries have formulated long term low greenhouse gas emission development strategies related to the Paris Agreement which aimed to meet the essential strategies addressing issues with climate change, environmental protection and low carbon.

The astonishing developments on unmanned aerial vehicle (UAV) systems and artificial intelligence (AI) technologies enables a great opportunity to monitor the environment and propose reliable solutions to restore and preserve the planet and human health.

Data acquisition and processing paradigm has been changed as a result of technological developments. It is obvious that new solutions, innovative approaches will make significant contributions to solve the problems which our planet is facing. UAV data can be collected by various platforms (planes or helicopters, fixed wing systems, drones) and sensors for earth observation and sustainable environmental monitoring which are also utilized by the United Nations to support the delivery of its mandates, resolutions, and activities.

UAV based earth observation data and AI techniques have a wide range of applications such as risk management, disaster monitoring and assessment, environmental impact evaluation and restoration, monitoring agriculture and food cycles, urban analysis, digital twin and smart city applications and providing increased situation awareness. This growth of widely available UAV data associated with the exponential increase in digital computing power, machine learning and artificial intelligence plays a key role in the environmental monitoring and solution generation of geospatial information for the benefit of humans and the planet.

The proposed special issue aims to contributes ASPRS's key mission on 'Simplify and promote the use of image-based geospatial technologies for the end-user', 'Promote collaboration between end users and geospatial experts to match data and technology to applications and solutions' and 'promote the transfer of geospatial data and information technology to developing nations' by serving as an innovative knowledge exchange platform for authors from the globe to deliberate on the latest advancements, state-of-the-art developments and solutions that can help the community to solve many real-world challenges on the topic of "*AI-Based Environmental Monitoring with UAV Systems*."

This special issue aims to bring researchers to share knowledge and their expertise about state-of-art developments and contribute to the goal of a livable world by integrating human creativity with UAV and AI technologies for environmental monitoring to combat global threats on ecosystems. We wish to discuss the latest developments, opportunities and challenges that can solve many real-world challenges in environmental monitoring including but not limited to:

- AI-Based UAV and GIS Applications
- AI-Based Object Detection and Recognition from UAV Imagery
- AI-Based Digital Twin Applications
- AI-Based Smart City Applications

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/.

Deadline for Manuscript Submission
May 1, 2022

Submit your Manuscript to
http://asprs-pers.edmgr.com

Guest Editor
Tolga Bakirman, PhD, Y*ildiz Technical University*, Turkey
Dr. Tolga Bakirman. bakirman@yildiz.edu.tr. is an assistant professor at Yildiz Technical University in the Department of Geomatic Engineering.

Remote Sensing and Human Factors Research: A Review

Raechel A. Portelli and Paul Pope

Abstract

Human experts are integral to the success of computational earth observation. They perform various visual decision-making tasks, from selecting data and training machine-learning algorithms to interpreting accuracy and credibility. Research concerning the various human factors which affect performance has a long history within the fields of earth observation and the military. Shifts in the analytical environment from analog to digital workspaces necessitate continued research, focusing on human-in-the-loop processing. This article reviews the history of human-factors research within the field of remote sensing and suggests a framework for refocusing the discipline's efforts to understand the role that humans play in earth observation.

Introduction

Understanding the earth's physical and cultural environments is critical for facing today's biggest challenges, such as climate change (Reichstein *et al.* 2019) and food security (Wen *et al.* 2021). Unfortunately, data collection currently outpaces knowledge generation (Maxwell *et al.* 2018). Therefore, remote sensing is becoming more reliant on artificial intelligence (AI) to facilitate different components of the image-analysis pipeline. However, even though these increasingly computer-based methods have improved the accuracy of data products for a range of applications—including agriculture, environmental epidemiology (VoPham *et al.* 2018), and sustainable development (Holloway and Mengersen 2018)—AI systems are notoriously brittle when presented with novel cases (Tuia *et al.* 2016).

Understanding human performance and barriers to success in interpretation has essential implications for the further development of geospatial AI. First, for as long as computational methods have been in development, human operator performance has served as a benchmark for computational image-analysis algorithms. Second, integrating human knowledge and domain expertise leads to improved outcomes in terms of increased reliability and robustness. Third, humans have cognitive abilities, such as contextualization, that scientists have yet to replicate adequately in a computational methodology. For these reasons, the research community must continue to be inclusive of human-oriented research.

The concept of human-extended machine cognition suggests a system that integrates human agents' activity with a collection of machine-based processing routines with the intent of leveraging the strengths of the two cognitive systems (Smart 2018). Human cognition is efficient and flexible when faced with visual tasks (Crowe 1998). The field of visual analytics has rallied around the idea that interactive visual interfaces enhance human visual interpretation, supporting discovery, hypothesis making, testing, and verification (Yusoff and Salim 2020). Geographic information scientists have proposed these types of systems in the past, but adoption by scientists in earth-observation research has been limited. Despite this, human-factors research concerning image interpretation is still carried out, although many studies are ancillary to a considerable amount of computationally focused scientific research.

We argue that the relegation of human-factors research to the fringe of earth-observation science is wholly at odds with other image-based sciences, thereby limiting possible innovation. However, science is supported by prior work, which suggests a need for a review of that body of work. Two previous research volumes addressing human factors relating to image analysis were published by Hoffman and Markman (2001) and White *et al.* (2018). Both volumes contain research perspectives and reports on recent expertise and perceptual-skill research from psychologists, remote sensing professionals, and geographic scientists. The earlier volume leans strongly toward terrain analysis, whereas the second volume introduces more cartographically oriented research. Many of the studies reported in those volumes are represented here through their peer-review publications.

This article reviews cognitive research in earth-observation remote sensing, its impact on computational methods of image analysis, and a conceptual model of the current research landscape. It presents the critical challenges that a cognitively informed approach to earth observation faces and the research questions addressing those challenges.

Background

In remote sensing image analysis, cognitive research is rooted in air photointerpretation training (Bianchetti and MacEachren 2015). The *Stanford Encyclopedia of Philosophy* defines *cognitive science* as the "interdisciplinary study of mind and intelligence, embracing philosophy, psychology, artificial intelligence, neuroscience, linguistics, and anthropology" (Thagard 2019). Cognitive science is an essential component of understanding geospatial data (Raubal 2018). Cognitive geographic information science is motivated by improving the usability, efficiency, equity, and profitability of geospatial data, and cuts across numerous geospatial applications (Montello 2009).

The earliest consideration of human factors of image interpretation informed the training of new interpreters during World War I (Campbell 2008). There was a need to select and train the best candidates. Tools such as the Army Individual Test of General Ability were established to evaluate recruits' prior experience and individual qualities. While air photointerpretation was relatively new at the onset of World War I,

Raechel A. Portelli is with Michigan State University, East Lansing, MI 48823 (raechel@msu.edu).

Paul Pope is with the Los Alamos National Laboratory, Los Alamos, NM 87545 (papope@lanl.gov).

Contributed by Dongdong Wang, March 12, 2021 (sent for review June 30, 2021; reviewed by Jessica Wardlaw).

Photogrammetric Engineering & Remote Sensing
Vol. 88, No. 1, January 2022, pp. 55–64.
0099-1112/22/55–64

doi: 10.14358/PERS.21-00012R2

there was little doubt regarding its importance by the end of World War II.

Several symposia were held throughout the 1960s in response to the technological innovations occurring in remote sensing. Human-factors research continued in government agencies such as the Army Personnel Research Office and the Human Factors Research Branch of the Adjutant General's Research and Development Command (Sadacca 1963). Verstappen (1962–1964) notes that the Delft Symposium in 1962 focused on human factors, especially activities of the USA and USSR. These types of human-performance studies continued throughout the 1970s, even after computational methods were developed. Much of this work is documented through government reports (e.g., Jeffrey *et al.* 1980). To provide context for this later work, we note that in 1960 the American Society of Photogrammetry (now the American Society for Photogrammetry and Remote Sensing) published the first edition of the *Manual of Photographic Interpretation.* The third chapter, by Rabben *et al.* (1959), outlines some of the human-factors research of the period.

In the United States, the threat of armed conflict during the Cold War period led to the rapid development of novel remote sensing technologies and demands for new interpretive techniques. Interpreter expertise continued to be a central concern for military development and research. Both Summerson (1954) and Schatzley and Karably (1954) provide general reviews of the different human-factors aspects of photointerpretation practice. At that time, research emphasized improving human abilities to overcome image-quality limitations (Martinek and Sadacca 1961). These studies addressed human-performance factors such as completeness, completion time, and accuracy, and image properties such as resolution, image type (e.g., natural color, infrared, side-looking airborne radar), presentation rate, and spatial scale. Symposia such as the Discussion on Man-Machine Interface in Photogrammetry, held 7 and 8 August, 1972, drew attention to the growing importance of more symbiotic relationships between interpreters and their tools and technologies for interpretation.

By the 1980s, expert or knowledge-based image-analysis systems emerged. Human-factors research took on a new role in systems development, including acquisition, representation, and use of domain knowledge held by experts (Estes *et al.* 1988). The general architecture consisted of an expert-knowledge database and a reasoner (or inference machine) developed through interactions with subject-matter experts. Such systems rely on users' heuristics (such as if-then statements) to guide the analysis process. However, as Noble (1987) points out, many of these systems were limited in their applicability to diagnostic problems. Despite these limitations, several expert systems were designed to augment image analysis during this decade and into the early 2000s (see, e.g., Goodenough *et al.* 1987; Nicolin and Gabler 1987; Sarma and Sarma 1994; McKeown *et al.* 1989; Prasad *et al.* 2002). Expert systems persist as stand-alone items, such as KIM (Datcu and Seidel 2005) and Tetracorder (Livo and Clark 2014), or as components of Geographic object-based image-analysis (GEOBIA) systems.

GEOBIA reignited the interest in human image interpretation and its role in rule-based image classification (Chen *et al.* 2018). Tools such as Trimble's eCognition can implement rule-based classification and integrate structured knowledge through XML schemas. Increased interest in ontology adoption has spurred research into encoding formalized expert knowledge (Rajbhandari *et al.* 2017). Integrating ontologies into image-analysis workflows helps bridge the semantic gap between low-level image content and human conceptualization (Durand *et al.* 2007).

In addition to these expert-oriented developments, volunteered geographic information and crowdsourcing for informing image classification have resulted in implementing human-factors-style research designs to better understand novice user training and the design of more complex collaborative classification systems (See *et al.* 2013; Wardlaw *et al.* 2018).

The past century of human-factors research in remote sensing is rich, yet it remains underdeveloped within the earth-observation community. Early- to mid-20th-century developments in understanding individual differences and training needs are recorded in gray literature that is not always readily accessible, such as military training manuals and reports. These would eventually influence the first conferences on aerial photointerpretation and thereby the first editions of various aerial photointerpretation books. Additionally, as computational analysis became more accessible and famous, it has become difficult to easily differentiate human-oriented and computationally oriented research as human-factors research, as terms are used interchangeably to describe computational and manual image interpretation. For example, the term *perception* is used extensively in both cases, despite fundamental differences in how computers and humans "see" an image. The remainder of this article presents a review of significant research thrusts in human factors of remote sensing.

Conceptual Framework

The compartmentalization of perception and cognition is debatable (for a review of this debate, see Montemayor and Haladjian 2017). For this article, we distinguish between perception and cognition as organizing concepts. Perception, namely visual perception, is the process of actively receiving input from the environment via a sense organ, in our case the eyes. *Visual perception* allows for the capture and processing of information. On the other hand, cognition is a higher-level process where knowledge and understanding are acquired through the senses, experience, or thought. In addition to visual perception and cognition, *knowledge representation* is used in this review to incorporate research on knowledge management and image analysis. This third category encapsulates the connection between what is perpetually encountered (*visual recognition*) and what activates the mind (*visual imagery*).

Image interpretation is classically defined as "the act of examining photographic images of objects for the purpose of identifying the objects and deducing their significance" (Colwell 1954). Johnson (1958) further refines the interpretation process in his framework, suggesting four levels of discrimination (Table 1). These four levels have been used extensively to study resolution requirements for techno-visual systems and offer complete definitions directly relatable to image interpretation. They are similar to Vink's (1964) tasks associated with interpretation: detection, recognition and identification, analysis, deduction, classification, and idealization. In a more recent assessment of image-interpretation tasks, Campbell (2011) suggests classification, detection, recognition, identification, enumeration, measurement, and delineation.

The interpretation processes in Table 1 have been used extensively to structure experimental approaches. Physical models were an early form of visual stimuli for the study of interpretation. While not directly studying image interpretation, these experiments informed image-acquisition practices

Table 1. Johnson's levels of discrimination.

Level	Meaning
Detection	An object is present.
Orientation	An object is approximately symmetrical or unsymmetrical, and its orientation may be discerned.
Recognition	The class to which the object belongs may be discerned.
Identification	The target can be discerned to the limit of the observer's knowledge.

and an understanding of visual search for aerial surveillance. For example, Duntley (1953) used model-size objects of interest (toy soldiers and military equipment) and moving automobiles to evaluate the relationships between visual contrast, aircraft speed, and height, and found that object-recognition success varied with visual search and recognition methods. The highest accuracy was achieved by searching strategically through the entire scene while intermittently returning to the search target. Such results can be readily compared to image-based studies during this same period.

Image-interpretation experiment design incorporates three critical elements, as shown in Table 2. First is the *interpreter* (or *user*). Interpreter traits are those attributes that make the individual unique among other interpreters, such as gender, age, and education. Many of the studies reviewed here examine the influence of individual differences such as expertise and gender on interpretation outcomes. Second, interpretation *tasks* are the mental processes that scientists are testing. They often focus on the general task of object recognition; however, specific cognitive functions related to that task can be measured (quantified). For example, *visual search* is the process of scanning an image's contents with the objective of detecting a target. *Visual attention* is the process of filtering relevant information from a given image. These two concepts are tightly coupled, as visual attention affects visual-search performance.

The third critical component of visual-recognition experiments is the *image* used. Images can be varied in many ways. First, they can be described based on the representation type (modality), such as false-color or natural-color imagery. Second, an image's *spatial resolution* is essential when considering the resolvability of objects in a scene. The concept of *extent* relates to resolution, functioning as the image's resolution and map scale. Finally, an image's *content* relates to the type of land surface visible in it.

In some cases the content might be individual objects, and in other cases it might be more abstract, informational patterns. Finally, the angle of image collection, or the subject's view angle, is described as orientation or *viewpoint*. Traditionally the terms *oblique* (high or low) and *nadir* are used; however, specific view angles may also be used to describe this element (characteristic).

Visual Task-Based Studies

A collection of articles representing experiments on image recognition using earth-observation data are presented next. Most represent research conducted by geographers and domain experts, although cognitive scientists' involvement is evident in many of them. Table 2 contains the three elements used to organize the review.

Table 2. Conceptual framework of human-factors research in remote sensing image interpretation.

Element	Traits	Measurement
Interpreter/ user	Expertise, domain, demographic information	Survey, interview
Image	Type, spatial resolution, extent, content, viewpoint	Segmentation, classification, band math
Tasks	Visual search, route planning, classification, digitization	Eye-tracking, self-report, accuracy, completion time, confidence, EEG, emotion detection

Interpreter (User)

The general goals of image-interpretation experiments are to improve the user experience, improve accuracy and precision, and understand core cognitive processes related to visual perception. User training, experience, age, and gender are all commonly used to describe research participants. While children as young as four years old have been included in this type of research, most studies have involved adults in the workforce or university. Studies have shown that children as young as four years old can identify familiar places from maps, aerial photographs, and even false-color images (Dale 1971; Svatonova and Rybansky 2014).

Expertise is often equated with training and experience. The inclusion of training or experience as an independent variable in these studies may be present via the number of years of training an individual has had, the type of training they have taken part in, or the specific domain they are trained in. For example, Martinek and Sadacca (1961) included officers and enlisted men from the US Army for testing. However, in academic studies, students are more often used as participants, sometimes trained in the field of geographic information science or other remote sensing domains.

University students are often used in the studies presented because access to experts within the workforce is often limited by the costs of diverting their attention away from their everyday work tasks. Šikl *et al.* (2019) included the largest expert group of the studies presented here (n = 39), as well as geography and psychology students. Experts outperformed students on average (approximately 80%; untrained viewers were slightly above 70%), and performance was slightly positively correlated with the expert group's experience level. Those results were supported by previous results of Loschky *et al.* (2015) suggesting that orientation affects aerial photointerpretation.

Image

One of the earliest concerns of human-factors research was the effect of view angle on user interpretation. It was generally thought that imagery captured from directly above (i.e., a nadir viewpoint) would be more difficult to interpret than oblique imagery (Bianchetti and MacEachren 2015). Today, viewpoint is more often considered as a difference between ground-based perception and nadir perception. For example, Loschky *et al.* (2015) used similar rapid categorization and found that terrestrial-view scene-gist recognition—an action in which the scene's semantic content is perceived—was viewpoint dependent but aerial-view scene recognition was viewpoint independent, and suggested that this is due to our natural gravitational frame.

Aerial images and high-resolution satellite images are the most-used stimuli for interpretation studies. Although images are sourced from various places, they generally have enough spatial resolution to identify single target objects. However, there is little cohesiveness across these studies in the description of the image stimuli. We suggest a more consistent and systematic description, as the term *aerial image* encompasses many different sensors and resolutions. For example, two studies used Landsat and MODIS imagery with 30-m and 250-m resolutions in interpretation tests (Svatonova and Rybansky 2014; Svatonova 2015). However, in both cases, the resolution was not explicitly addressed as a potential factor that influenced interpretation.

In addition to resolution, consideration of image color representation is important. Not all sensors are created to be spectrally equivalent. While it may not be essential for image interpreters to be familiar with the specific spectral bandwidths of a given image, spectral combinations influence image patterns. In general, natural-color imagery (Figure 1a) or black-and-white imagery (Figure 1b) is used for visual research studies. Less common is false-color imagery (Figure 1c), although several studies by Svatonova have included this modality. There are, of course, other combinations and indices that use band math to convey spectral information. In Figure 1d, we provide an example of the normalized difference vegetation index. Unfortunately, there is minimal research on the effect of these other visual representations. One example that does include alternative representations is from Bianchetti (2016). Landsat Thematic Mapper imagery was presented as the visual result of a spectral transformation leading to very bright green and pink images (Figure 1d). However, this study's focus was on conceptualization and task handling, not visual response. Finally, synthetic aperture radar (SAR) was used in a target-detection experiment (Matzen *et al.* 2016). Experts familiar with SAR imagery performed faster and more accurately than those who were unfamiliar with SAR. Nonexperts more easily identified more salient targets, but saliency did not affect experts' metrics.

Another consideration is the real-world environment that an image represents. In some cases, like Saralioglu and Gungor (2019), the built-up environment is equivalent to residential space, and in other cases, dense urban scenes are used (Sowden *et al.* 1996). The selection of images for these studies is study-specific, and reanalysis of the same image set is not typical. An exception to this is the series of images used in scene-gist studies (Pannasch *et al.* 2014; Loschky *et al.* 2015). It is also interesting to note that few studies consider uninhabited environments. One notable exception is work considering the visual interpretation of geomorphology on the Martian surface (Wardlaw *et al.* 2018). This focus on inhabited space is for several reasons. First, high-resolution images of built environments are likely to be the most familiar images encountered by people. They are often associated with navigational web portals like Google Maps or Bing Maps. Second, the participants themselves are more likely to directly interact with these types of environments in their daily lives and, in return, become more familiar with them than, say, an environment devoid of human activity. Finally, it is easier to specify objects of interest in built environments than in vast uninhabited spaces, making the image-interpretation task easier for the participants and the accompanying analysis of experimental results more manageable for the scientists.

Task

As already discussed, several frameworks suggest potential organizations of image-interpreter tasks. Most studies surveyed in this review address visual search and classification. Visual search in this context is the process of visually scanning an image to detect some target. This target is often an object, for example, a familiar location (Dale 1971) or even a UFO (Zelinsky and Schmidt 2009). In other cases, participants are tasked with locating regions or areas of interest.

Visual saliency (the perceptual quality of an object to stick out in comparison to its surroundings) affects visual search. However, it is most likely not the only controlling factor in visual attention. Henderson (2003) notes that knowledge and familiarity may serve as top-down influences on visual attention. Experimental results have previously shown that how experts structure their domain knowledge about the environment affects their categorization of imagery (Medin *et al.* 1997).

Only one of the studies identified as part of this review considered digitization, which is surprising given the extensive efforts to improve boundary-accuracy measurement (Radoux and Bogaert 2017). Van Coillie *et al.* (2014) performed a set of analyses examining individual operator effects, including demographic characteristics and personality factors. Participants digitized a series of objects from remote sensing imagery with a variety of geometric and spectral properties. Digitizations were evaluated for completeness, correctness, accuracy, and positional accuracy. Not surprisingly, operator digitizations varied considerably, which was attributed to both individual differences and variations in the target object.

Another type of experiment—route-planning tasks—requires a participant to trace a path from one point to another based on a given criterion. Two studies have assessed the results of the route-planning task when assigned to children. Sowden *et al.* (1996) had child participants aged four years use a photograph to plot out a route between two locations and identify a set of features in the landscape. Fourteen of the 20 children developed appropriate navigation routes, and 13 of the 20 made no errors in identifying features. Svatonova and Rybansky (2014) asked students (ages 11, 15, and 19) to create a shortest-path route between two points as seen on imagery, in an experiment to test the influence of different map backgrounds. Older participants performed more accurately than younger ones in both map and image cases, with 15-year-olds outperforming both 11-year-olds and 19-year-olds in the route-planning task.

The classic tasks of land cover and land use classification are found in many earth-observation research articles. A variety of classification schemas are used to organize our knowledge about the environment (see the "Conceptualization Studies" section), but land use and land cover classification are pervasive. Thus, it is no surprise that many human-factors experiments have addressed this topic. For example, Lloyd and Hodgson (2002) performed a series of classification experiments based on land cover and land use categories defined by Anderson *et al.* (1976). They found that image complexity impeded visual search, but objects unique to specific land uses could reduce visual search time. Earlier, Hodgson (1998) had evaluated the effect of window size on land use classification and found that increased window size improved accuracy, but that this improvement leveled off at a window size of about 40×40 pixels, except in the case of commercial land use.

Sometimes classification is based on visual estimation. For example, Battersby and Hodgson (2012) found that experts more accurately estimated hurricane damage from

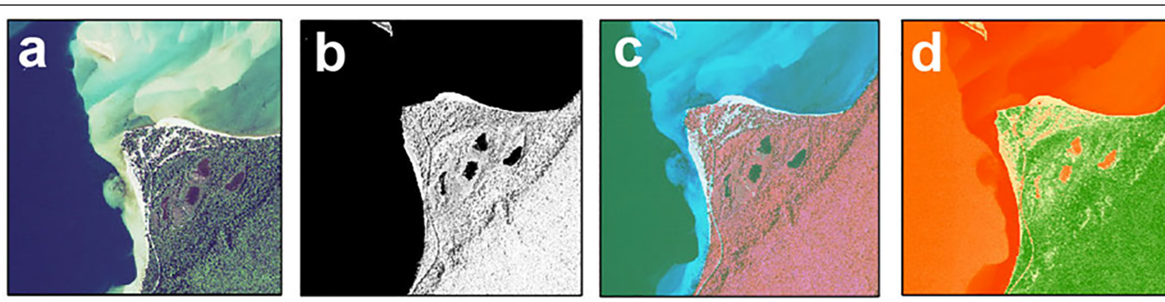


Figure 1. Examples of image types used in remote sensing visual-interpretation research: (a) natural-color imagery, (b) black-and-white imagery, (c) false-color imagery, (d) derivative imagery.

high-resolution imagery (86.2% for image experts, 86.4% for hazards experts, and 81.9% for novices). This estimation task is commonly found in crowdsourcing tasks (e.g., Barrington *et al.* 2011). However, estimation may also be based on more abstract image qualities, such as tonal differences. For example, Cihlar and Protz (1972) found that in participants' ranking of tonal differences in black-and-white imagery, there was a linear relationship between these perceptual tonal differences and computational color difference formulas.

The process of change detection requires the perception of some alteration to an image pattern. Davies *et al.* (2006) found that experts were more responsive to visual saliency (as quantified by the method of Itti and Koch, 2001) than nonexperts in change-detection experiments. In a follow-up study, Lansdale *et al.* (2010) introduced the use of eye-tracking to refine their ability to detect effects of visual saliency. In this second study, a significant relationship between salience and fixation by nonexperts was found, but no significant relationship was found for their expert group.

Wardlaw *et al.* (2018) presented high-resolution images of the Martian environment to conduct a change-detection experiment among doctoral students and postdoctoral researchers with expertise in planetary science. These participants were presented with the task of identifying and marking change, and attributing the type of change that occurred, in the black-and-white images. Consensus analysis was carried out to examine differences between interpretations. Low levels of agreement were found within the expert group and between the expert and nonexpert groups.

Target detection is a common motivation for interpretation research. In these studies, participants are prompted to find a target in a scene. The following studies examine viewpoint, participant familiarity, context, and the effect of image enhancements. To summarize their findings, target detection is more efficient and effective when a ground-based viewpoint is used (Martinek and Sadacca 1961; Pannasch *et al.* 2014), when scene context is provided during the prompt (Zelinsky and Schmidt 2009), and when enhancement is used to improve the visual saliency of targets that are particularly difficult to perceive (Dong *et al.* 2014). Perhaps surprisingly, there does not seem to be an effect of the type of image representation (false or natural color) on target detection.

Martinek and Sadacca (1961) found that neither viewpoint (oblique or nadir) nor display (monoscopic or stereoscopic) resulted in a statistically significant difference in correctness on the task of truck-convoy detection. Pannasch *et al.* (2014) found that duration and saccadic amplitude varied with orientation and viewpoint (terrestrial versus aerial and inverted orientation versus regular). An aerial viewpoint led to a longer time to the first fixation, and fixation duration and saccade amplitude were also prolonged. They also found that the similarity of the visual scan path between participants was higher when comparing terrestrial scenes than aerial scenes, suggesting that familiarity may play a role in the search strategy.

Dale (1971) compared recognition of familiar places from maps and aerial photographs in children between the ages of 7 and 11 from a village in the UK. Children were interviewed in pairs to understand whether the use of maps versus images was more intuitive for determining village locations. They preferred aerial photographs when they were less familiar with the neighborhood, but preferred a map when they were more familiar with the neighborhood.

Zelinsky and Schmidt (2009) implemented a targeted search for a UFO in aerial photographs under conditions with and without region-context information. Participants were presented with the target cue and asked to locate the target. The results indicate that the contextual cues led to faster target detection with fewer fixations. Additionally, participants' gazes were constrained by the region context provided. For example, if the target cue was "road," the participants' gaze was more prominent on the image's road infrastructure. This contextual benefit was found to be more beneficial in some land cover types than in others.

Svatonova (2015) found faster and more accurate target detection with the use of aerial photographs compared to topographic maps for emergency management. The results also indicate that target detection was not affected by whether the display was false or natural color.

Dong *et al.* (2014) evaluated the effects of several image enhancements on user identification of regions of interest. Their results indicate that enhancement improved detection, especially when the target features were less visually salient in the original image due to the environment (e.g., obstruction of roads due to trees). Table 3 summarizes these and other relevant visual-interpretation studies.

Crowdsourcing (an alternative to individual interpretation) has become a standard method for incorporating image interpretation into image analysis. The power of crowdsourcing lies in the assumption that group consensus outperforms individual expert performance. For example, Saralioglu and Gungor (2019) found that classification tasks achieved higher accuracy when majority-vote crowdsourcing was implemented than when a single expert was used to validate the algorithmic output. Their results corroborate what Albuquerque *et al.* (2016) found: 99% accuracy achieved based on consensus voting.

Albuquerque *et al.* (2016) present a typology of crowdsourcing tasks for generating geospatial information: classification, or the recognition of objects and assignment of an additional attribute; digitization, or the creation of new geographic objects; and conflation, or the combination of geographic information to improve it in some way, such as fusion of various sources or redundancy reduction. They go on to present a case study based on identifying settlements, roads, or waterways. Their results suggest that images containing such features were more frequently misclassified. They attribute this misclassification to participants overlooking the presence of crucial indicator objects, since images without these objects were more accurately classified.

Conceptualization Studies

Knowledge representation is the process of modeling real-world knowledge by using computer data structures. Conceptual definition and organization were central components to the design of expert systems, and they continue to be essential for the development of interoperable geospatial information (Kuhn 2005). A critical difference between novice and expert interpretation capabilities is inference (Gegenfurtner and van Merriënboer 2017). While both novices and experts can readily note the presence or absence of features in an image, experts have additional domain knowledge that allows them to make inferential conclusions that nonexperts cannot. For example, visual evidence of environmental changes, like drought, requires expertise about climatic, vegetation, and landform patterns within the context of a given landscape. The visual identification of brown vegetation could indicate many environmental conditions; however, an expert will bring additional knowledge to provide context to the situation. This contextual knowledge is integral to experts' attribution of land cover change agents (Bianchetti 2016).

One goal for human-factors research is to extract, organize, and translate expert conceptualizations into computer-readable formats that can be integrated with computational methods.

Ontologies are explicit, formal specifications of domain concepts and the relations among them (Gruber 1995). They provide several key benefits to the remote sensing process,

Table 3. Studies concerning the human factors of visual interpretation.

Reference	Topic	Task	Method	Domain	Stimuli
Matzen *et al.* (2016)	Visual saliency	Target detection	Eye-tracking	Terrain features	SAR
Lansdale *et al.* (2010)	Visual saliency	Change detection	Eye-tracking	Urban/rural	Aerial photography
Lansdale *et al.* (2010)	Short-term visual memory	Change detection	Recall	Urban/rural	Aerial Photography
Lloyd and Hodgson (2002)	Categorization	Classification	Recall	Level 1/Level II ALULC	Aerial photography
Battersby and Hodgson (2012)	Categorization	Ranking	Classification	Damage assessment	Aerial photography
Svatonova (2015)	Expertise	Target detection	Detection	Flood and fire threats	Landsat (color and false color), aerial photography, topographic map
Cöltekin *et al.* (2017)	Memory	Route planning	Recall	Route planning	Orthophoto map versus map
Sowden *et al.* (1996)	Map reading	Navigation and feature detection	Observation	Urban	Black-and-white photography
Svatonova and Rybansky (2014)	Visual memory	Object detection	Route planning	Urban	Landsat and aerial photography (color)
Šikl *et al.* (2019)	Expertise	Recognition/memory	Recall	Urban	Vertical and oblique aerial imagery
Dale (1971)	Visual memory	Object detection	Recall	Village features	Color photography
Hodgson (1998)	Categorization	Categorization	Classification	Land use	Aerial photography
Van Coillie *et al.* (2014); Gardin *et al.* (2011)	Personality	Digitization	Multiple	Urban scenes	Aerial photography
Loschky *et al.* (2015)	Scene gist	Categorization	Classification	Natural versus constructed	Aerial photography
Cihlar and Protz (1972)	Tonal discrimination	Classification	Rating/rank difference	Agriculture	Black-and-white aerial photography
Pannasch *et al.* (2014)	Visual search	Categorization	Eye-tracking	Urban UFO	Aerial photography
Saralioglu and Gungor (2019)	Validation	Crowdsourcing	Object categorization	Residential area	WorldView-2 and pan-sharpened WV2
Albuquerque *et al.* (2016)	Validation	Crowdsourcing	Feature detection (Y/N)	Road or settlement/ Presence	Bing imagery
See *et al.* (2013)	Validation	Crowdsourcing	Classification	Land cover and human impact	Google imagery
Wardlaw *et al.* (2018)	Expertise	Crowdsourcing	Change detection	Geomorphology	Martian imagery

Table 4. Studies concerning the organization of conceptual information.

Reference	Goal	Method	Domain	Imagery
Hoffman (1984, 1995)	Expertise	Terrain feature classification	Terrain	Aerial photography and radar
Hoffman (1991)	Expertise	Observation	Weather forecasting	GOES
Pope (2012)	Ontology-Dev	Annotation	NAICS	Aerial photography
Hoffman and Pike (1995)	Categorization	Interviews	Classification	
Bianchetti (2016)	Expertise	Observation and interview	LULC change	Landsat TM and aerial photographs (BW)
Kohli *et al.* (2012)	Ontology-Dev	Interviews, semi-structured surveys	Slum mapping	GeoEye-1
Aryal and Josselin (2012)	Segmentation interpretation	Description	Land cover	Google Earth

including their multimodal representation ability, cognitive semantic reasoning capabilities, and interoperability (Arvor *et al.* 2019). As Arvor *et al.* point out, knowledge-driven analyses require the conversion of conceptual models held within experts' minds into numeric models through their understanding of remote sensing principles. The extraction of these conceptual models requires the use of knowledge-elicitation techniques. While numerous studies have been published regarding the development of automated knowledge-based systems from expert-knowledge databases, most of do not document the process of eliciting expert knowledge in the construction of those databases. This section reviews those that do.

The introduction of ancillary information is essential for guiding the image-interpretation process as performed by experts. For example, Martinek and Sadacca (1961) found that participants improved their target-detection rates when intelligence information was provided in cases where the visual perception of targets was inhibited or ambiguous. Additionally, photointerpretation keys, especially dichotomous keys, were used to document and standardize decision making during visual interpretation.

The early development of expert systems relied on interpreters to parse rules generated from experts to generate decisions (Mulder 1985). Hoffman (1984, 1995) used interviews and observations to develop preliminary information about expert terrain interpretation using aerial photographs and radar. This study was followed by a study linking generic terrain definitions to symbolic representation (Hoffman 1985). In another set of studies, Hoffman (1991) used observation to develop a set of design recommendations for developing a meteorological forecasting system.

Bianchetti (2016) used a mixture of knowledge-elicitation techniques to evaluate the classification strategies of expert image interpreters performing change-detection and attribution tasks. These experts used false-color representations and high-resolution aerial imagery to understand the reasoning behind land cover change interpretation. The results indicate that the level of expertise did not influence the outcome of the interpretation; however, training history and experience did influence the reasoning methods used to arrive at the final interpretation.

Studies incorporating human interpretation in knowledge formalization for ontology development include that of Kohli *et al.* (2012), who used a combination of literature review, interviews, and a semi-structured survey of 51 experts to develop an ontology for slum identification from imagery, called the Generic Slum Ontology. Pope (2012) conducted an annotation experiment with experts to evaluate consensus between annotators in industrial-site identification. In other instances, semantic information has been extracted from previously published results and existing classifications (Rajbhandari *et al.* 2017) as well as small groups of experts tasked with direct interpretations (Belgiu *et al.* 2014).

A small study conducted by Aryal and Josselin (2012) examined the identification of land cover classes from a segmented aerial photograph. This study's sample size was too small to be statistically examined, but its findings indicate potential differences in semantic interpretation based on expertise type, specifically between ecologists and geographers. A second task, the location of specific objects from the image, was used to examine scale differences related to object detection. While the research is inconclusive, it points to additional dimensions that could be tested in future studies.

Arvor *et al.* (2013) describe several challenges to integrating ontologies with remote sensing analyses. First, there is a need to integrate and understand the role of vagueness and uncertainty in the segmentation of real-world objects. Second, integrating qualitative information about spatial features with quantitative image data must be refined and systematized. Third, many features are not visible or applicable across all scales, so there is a need to develop multi-scalar ontologies in practice. Fourth, features changing through time are not captured in static ontologies, so their capture is needed. Finally, fundamental assumptions made in remote sensing analysis must be considered, most notably the discrepancy between "open-world" conceptualizations of experts and "closed-world" assumptions generally made during remote sensing image analysis.

Discussion and Conclusion

The 20th century saw the development of two approaches to extracting information from remotely sensed images: on the one hand, human interpretation based on the direct visual perception of analog images or digital displays; on the other, computational approaches using statistical and mathematical modeling. These two approaches have complementary strengths and weaknesses. Within other visually oriented domains, such as radiology, integrating these two systems has led to improved outcomes (Pesapane *et al.* 2020). To create these synthetic, analytical systems, both users and designers must understand what human experts can bring to the table.

Trends

Most of the research reviewed here addressed the use of natural-color aerial imagery. Given the ubiquity of this type of imagery in web mapping and popular culture, its importance cannot be overstated. While some studies, such as the one Matzen *et al.* (2016), used less-common imagery types, such as SAR, a great opportunity exists for expanding this aspect of remote sensing human-factors work.

Three types of tasks have dominated these works: categorization, identification, and ranking. These are all everyday tasks that support image analysis, whether for scientific or general purposes. The metrics used in assessing such tasks, including accuracy, completion time, and precision, have a strong foundation in traditional psychological research. In addition to these types of metrics, increasingly technological measurements, such as those gathered through eye-tracking and functional MRI, are being included in human-factors experimental designs.

A critical aspect of human factors in remote sensing research is the participants. Over the past century, it has become possible for nonexperts to become more involved in the execution of this research. As research shifted focus from military objectives, university participation in the 1970s and 1980s increased access to students and community members. Later, with the proliferation of the Internet came the possibility of global participation in some studies. This is particularly evident in the most recent studies using crowdsourcing.

One difficulty in the execution of human-factors research is lack of access to expert practitioners. As with other domains, expert time is costly and limited. In response, many scientists have used pseudo-expert students, potentially limiting the transferability of results into best practices to guide practical applications.

Another aspect of image interpretation that has not been adequately addressed in the studies reviewed here is collaborative sense making. While cartographic researchers have been studying this process for over two decades (for an early example, see MacEachren 2000), the remote sensing work reviewed here has primarily been viewed as a solitary process, except for the crowdsourcing studies.

Human–Machine Integration

The studies presented here cover the gamut of research concerning human interpretation and expertise for earth observation. We hope that this review is a starting point for reinvigorating human-factors research within the field. The viability of such integrative work is reflected in funding initiatives such as the National Science Foundation's Future of Work at the Human-Technology Frontier. Human-factors research in other domains (such as medical imaging) has led to improvements in training and performance outcomes (Nakashima *et al.* 2013). This research has informed the development of novel computational approaches to remote sensing analysis like Itti, Koch, and Niebur's computation model of visual perception (Li and Itti 2011).

One of the most explicit and perhaps successful uses of expert knowledge in remote sensing today is TimeSync, a time-series segmentation tool used to support large-scale computational forest change research (Cohen *et al.* 2008). TimeSync allows for the integration of expert knowledge about forest transition states to guide large-scale computational analyses. Its intentional design brings to bear domain-specific knowledge with broadly applicable image-understanding tasks to improve analytical outcomes.

GEOBIA

Human expertise has also been discussed concerning GEOBIA, an approach to image analysis that extends earlier work in expert-systems design. GEOBIA is often touted as incorporating human knowledge into the analytical process (Blaschke *et al.* 2014). Human knowledge is integrated into these systems via one of three processes: digitization of boundaries to improve segmentation, definition of classification schemes using explicit ontologies, or implementation of rule-based classification. However, the need to handle multi-spatial input (e.g., satellite, aerial, and drone), asynchronicity in data acquisition (multi-temporal), and the variance in semantics induced thereby (e.g., forest versus the individual trees and their leaves) is an area of research that is only just beginning to be addressed.

Visualization

Knowledge about remote sensing image interpretation can also inform the development of geo-visualization products, as in work performed by Hoarau and colleagues (Hoarau 2012; Hoarau *et al.* 2013) concerning imagery-based map design and work by Dong *et al.* (2014) concerning image enhancement for base map design. The evaluation of user color preferences conducted by Mirijovsky and Popelka (2016) indicates that perceptual tendencies, such as dwell time and fixation, represent user preference. Studies like these and others can go a long way toward improving the ability of earth-observation scientists to communicate both within their scientific domain and to laypersons. Cartographers have a long history of cognitive-science research that deserves the earth-observation community's attention beyond the standard advice to avoid rainbow color maps (Phipps and Rowe 2010).

This article presented a review of research addressing the human factors governing remote sensing image interpretation. We have summarized results from two different research bodies, namely visual task-based research and knowledge representation. In addition to this review, we have highlighted the key areas where we believe human-factors research can improve analytical outcomes: GEOBIA, crowdsourcing, and visualization. Finally, we hope that cognizance of past precedent will help avoid previous pitfalls and build upon known successes.

References

Albuquerque, J. P., B. Herfort and M. Eckle. 2016. The tasks of the crowd: A typology of tasks in geographic information crowdsourcing and a case study in humanitarian mapping. *Remote Sensing* 8(10):859. https://doi.org/10.3390/rs8100859.

Anderson, J. R., E. E. Hardy, J. T. Roach and R. E. Witmer. 1976. *A Land Use and Land Cover Classification System for Use with Remote Sensor Data. United States Geological Survey Professional Paper 964.* Washington, DC: US Government Printing Office. https://doi.org/10.3133/pp964.

Arvor, D., M. Belgiu, Z. Falomir, I. Mougenot and L. Durieux. 2019. Ontologies to interpret remote sensing images: Why do we need them? *GIScience & Remote Sensing* 56(6):911–939. https://doi.org/10.1080/15481603.2019.1587890.

Arvor, D., L. Durieux, S. Andrés and M.-A. Laporte. 2013. Advances in Geographic Object-Based Image Analysis with ontologies: A review of main contributions and limitations from a remote sensing perspective. *ISPRS Journal of Photogrammetry and Remote Sensing* 82:125–137. https://doi.org/10.1016/j.isprsjprs.2013.05.003.

Aryal, J. and D. Josselin. 2012. Association of spatial object recognition and GEOBIA: An experimental approach. Pages 19–24 in *Multidisciplinary Research on Geographical Information in Europe and Beyond: Proceedings of the AGILE'2012 International Conference on Geographic Information Science*, held in Avignon, France, 24–27 April 2012. Edited by J. Gensel, D. Josselin and D. Vandenbroucke.

Barrington, L., S. Ghosh, M. Greene, S. Har-Noy, J. Berger, S. Gill, A. Y.-M. Lin and C. Huyck. 2011. Crowdsourcing earthquake damage assessment using remote sensing imagery. *Annals of Geophysics* 54(6). https://doi.org/10.4401/ag-5324.

Battersby, S. E., M. E. Hodgson and J. Wang. 2012. Spatial resolution imagery requirements for identifying structure damage in a hurricane disaster. *Photogrammetric Engineering and Remote Sensing* 78(6):625–635. https://doi.org/10.14358/PERS.78.6.625.

Belgiu, M., L. Drăguăand J. Strobl. 2014. Quantitative evaluation of variations in rule-based classifications of land cover in urban neighbourhoods using WorldView-2 imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* 87:205–215. https://doi.org/10.1016/j.isprsjprs.2013.11.007.

Bianchetti, R. A. 2016. Describing the problem-solving strategies of expert image interpreters using graphical knowledge elicitation methods. *GIScience & Remote Sensing* 53(5):561–577. https://doi.org/10.1080/15481603.2016.1196424.

Bianchetti, R. A. and A. M. MacEachren. 2015. Cognitive themes emerging from air photo interpretation texts published to 1960. *ISPRS International Journal of Geo-Information* 4(2):551–574. https://doi.org/10.3390/ijgi4020551.

Blaschke, T., G. J. Hay, M. Kelly, S. Lang, P. Hofmann, E. Addink, R. Q. Feitosa, F. van der Meer, H. van der Werff, F. van Coillie and D. Tiede. 2014. Geographic Object-Based Image Analysis—towards a new paradigm. *ISPRS Journal of Photogrammetry and Remote Sensing* 87:180–191. https://doi.org/10.1016/j.isprsjprs.2013.09.014.

Campbell, J. B. 2008. Origins of aerial photographic interpretation, U.S. Army, 1916 to 1918. *Photogrammetric Engineering and Remote Sensing* 74(1):77–93. https://doi.org/10.14358/PERS.74.1.77.

Campbell, J. B. and R. H. Wynne. 2011. *Introduction to Remote Sensing*, 5th ed. New York: Guilford Press.

Chen, G., Q. Weng, G. J. Hay and Y. He. 2018. Geographic object-based image analysis (GEOBIA): Emerging trends and future opportunities. *GIScience & Remote Sensing* 55(2):159–182. https://doi.org/10.1080/15481603.2018.1426092.

Cihlar, J. and R. Protz. 1972. Perception of tone differences from film transparencies. *Photogrammetria* 28(4):131–140. https://doi.org/10.1016/0031-8663(72)90010-5.

Cohen, W. B., Z. Yang and R. E. Kennedy. 2008. TimeSync: Synching human and automated interpretations of Landsat time-series to support a new paradigm in land cover change detection. Page GC13A-0678 in *AGU Fall Meeting Abstracts*, held in San Francisco, California, 15–19 December 2008.

Colwell, R. N. 1954. A systematic analysis of some factors affecting photographic interpretation. *Photogrammetric Engineering* 20(3):433–454.

Crowe, S. F. 1998. The differential contribution of mental tracking, cognitive flexibility, visual search, and motor speed to performance on parts A and B of the Trail Making Test. *Journal of Clinical Psychology* 54(5):585–591. https://doi.org/10.1002/(SICI)1097-4679(199808)54:5<585::AID-JCLP4>3.0.CO;2-K.

Dale, P. F. 1971. Children's reactions to maps and aerial photographs. *Area* 3(3):170–177.

Datcu, M. and K. Seidel. 2005. Human-centered concepts for exploration and understanding of Earth observation images. *IEEE Transactions on Geoscience and Remote Sensing* 43(3):601–609.

Davies, C., W. Tompkinson, N. Donnelly, L. Gordon and K. Cave. 2006. Visual saliency as an aid to updating digital maps. *Computers in Human Behavior* 22(4):672–684. https://doi.org/10.1016/j.chb.2005.12.014.

Dong, W., H. Liao, R. E. Roth and S. Wang. 2014. Eye tracking to explore the potential of enhanced imagery basemaps in web mapping. *The Cartographic Journal* 51(4):313–329. https://doi.org/10.1179/1743277413Y.0000000071.

Duntley, S. Q. 1953. *The Limiting Capabilities of Unaided Human Vision in Aerial Reconnaissance*. Ann Arbor, MI: Armed Forces, National Research Council, Vision Committee Secretariat.

Durand, N., S. Derivaux, G. Forestier, C. Wemmert, P. Gancarski, O. Boussaid and A. Puissant. 2007. Ontology-based object recognition for remote sensing image interpretation." In *19th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2007)*, held in Patras, Greece, 29–31 October 2007. https://doi.org/10.1109/ICTAI.2007.111.

Estes, J. E., M. A. Friedl and J. L. Star. 1988. Advanced feature extraction in remote sensing using artificial intelligence and geographic information systems. *Proceedings of SPIE* 0924:297–304. https://doi.org/10.1117/12.945698.

Gardin, S., S.M.J. van Laere, F.M.B. van Coillie, F. Anseel, W. Duyck, R. de Wulf and L.P.C. Verbeke. 2011. Remote sensing meets psychology: A concept for operator performance assessment. *Remote Sensing Letters* 2(3):251–257.

Gegenfurtner, A. and J. van Merriënboer. 2017. Methodologies for studying visual expertise. *Frontline Learning Research* 5(3):1–13. https://doi.org/10.14786/flr.v5i3.316.

Goodenough, D. G., M. Goldberg, G. Plunkett and J. Zelek. 1987. An expert system for remote sensing. *IEEE Transaction on Geoscience and Remote Sensing* GE-25(3):349–359. https://doi.org/10.1109/TGRS.1987.289805.

Gruber, T. R. 1995. Toward principles for the design of ontologies used for knowledge sharing? *International Journal of Human-Computer Studies* 43(5–6):907–928. https://doi.org/10.1006/ijhc.1995.1081.

Henderson, J. M. 2003. Human gaze control during real-world scene perception. *Trends in Cognitive Sciences* 7(11):498–504. https://doi.org/10.1016/j.tics.2003.09.006.

Hoarau, C. 2012. Orthoimage or map visualization in use in geoportals, case study on the French Geoportal. Pages PP–PP in *Proceedings, AutoCarto 2012*, held in Columbus, Ohio, 16–18 September 2012. Edited by J. Editor. Location: Publisher.

Hoarau, C., S. Christophe and S. Mustière. 2013. Mixing, blending, merging or scrambling topographic maps and orthoimagery in geovisualization? Pages PP–PP in *Proceedings of 26th International Cartographic Conference (ICC'13)*, held in Dresden, Germany, Month DD YYYY. Edited by J. Editor. Location: Publisher.

Hodgson, M. E. 1998. What size window for image classification? A cognitive perspective. *Photogrammetric Engineering and Remote Sensing* 64(8):797–807.

Hoffman, R. R. 1984. *Methodological Preliminaries to the Development of an Expert System for Aerial Photo Interpretation, ETL-0342*. Fort Belvoir, Va.: US Army Corps of Engineers, Engineer Topographic Laboratories.

Hoffman, R. R. 1985. *What Is a Hill? An Analysis of the Meanings of Generic Topographic Terms, ETL-0576*. Fort Belvoir, Va.: US Army Corps of Engineers, Engineer Topographic Laboratories.

Hoffman, R. R. 1991. Human factors psychology in the support of forecasting: The design of advanced meteorological workstations. *Weather and Forecasting* 6(1):98–110.

Hoffman, R. R. and A. B. Markman, eds. 2001 *Interpreting Remote Sensing Imagery: Human Factors*. Boca Raton, Fla.: Lewis Publishers.

Hoffman, R. R. and R. J. Pike. 1995. On the specification of the information available for the perception and description of the natural terrain. In *Local Applications of the Ecological Approach to Human-Machine Systems*, edited by P. Hancock, J. Flach, J. Caird and K. Vicente, 285–323. Boca Raton, Fla.: CRC Press. https://doi.org/10.1201/9780203748749-10.

Holloway, J. and K. Mengersen. 2018. Statistical machine learning methods and remote sensing for sustainable development goals: A review. *Remote Sensing* 10(9):1365. https://doi.org/10.3390/rs10091365.

Itti, L. and C. Koch. 2001. Computational modelling of visual attention. *Nature Reviews Neuroscience* 2(3):194–203. https://doi.org/10.1038/35058500.

Jeffrey, T. E., H. Martinek, U. Shvern and E. M. Johnson. 1980. *ARI Image Interpretation Research: 1970-1980, Research Report 1252*. Alexandria, Va.: US Army Research Institute for the Behavioral and Social Sciences. <https://apps.dtic.mil/sti/citations/ADA095661> Accessed DD Month YYYY.

Kohli, D., R. Sliuzas, N. Kerle and A. Stein. 2012. An ontology of slums for image-based classification. *Computers, Environment and Urban Systems* 36(2):154–163.

Kuhn, W. 2005. Geospatial semantics: Why, of what, and how? In *Journal on Data Semantics III*. Lecture Notes in Computer Science vol. 3534. https://doi.org/10.1007/11496168_1.

Lansdale, M., G. Underwood and C. Davies. 2010. Something overlooked? How experts in change detection use visual saliency. *Applied Cognitive Psychology* 24(2):213–225. https://doi.org/10.1002/acp.1552.

Li, Z. and L. Itti. 2011. Saliency and gist features for target detection in satellite images. *IEEE Transactions on Image Processing* 20(7):2017–2029. https://doi.org/10.1109/TIP.2010.2099128.

Livo, K. E. and R. N. Clark. 2014. *The Tetracorder User Guide: Version 4.4, Open-File Report 2013–1300*. Reston, Va.: US Geological Survey. <https://doi.org/10.3133/ofr20131300> Accessed 22 November 2021.

Lloyd, R. and M. E. Hodgson. 2002. Visual search for land use objects in aerial photographs. *Cartography and Geographic Information Science* 29(1):3–15. https://doi.org/10.1559/152304002782064592.

Loschky, L. C., R. V. Ringer, K. Ellis and B. C. Hansen. 2015. Comparing rapid scene categorization of aerial and terrestrial views: A new perspective on scene gist. *Journal of Vision* 15:11. https://doi.org/10.1167/15.6.11.

MacEachren, A. M. 2000. Cartography and GIS: Facilitating collaboration. *Progress in Human Geography* 24(3):445–456. https://doi.org/10.1191/030913200701540528.

Martinek, H. and R. Sadacca. 1961. Human factors studies in image interpretation: Part I. Introduction. *Photogrammetric Engineering* 27(5):714–728.

Matzen, L. E., M. J. Haass, J. Tran and L. A. McNamara. 2016. Using eye tracking metrics and visual saliency maps to assess image utility. In *IS&T International Symposium on Electronic Imaging 2016, Human Vision and Electronic Imaging 2016*, held in San Francisco, Calif., 14–18 February 2016. https://doi.org/10.2352/ISSN.2470-1173.2016.16.HVEI-127.

Maxwell, A. E., T. A. Warner and F. Fang. 2018. Implementation of machine-learning classification in remote sensing: An applied review. *International Journal of Remote Sensing* 39(9):2784–2817. https://doi.org/10.1080/01431161.2018.1433343.

McKeown Jr, D. M., W. A. Harvey and L. E. Wixson. 1989. Automating knowledge acquisition for aerial image interpretation. *Computer Vision, Graphics, and Image Processing* 46(1):37–81. https://doi.org/10.1016/S0734-189X(89)80016-7.

Medin, D. L., E. B. Lynch, J. D. Coley and S. Atran. 1997. Categorization and reasoning among tree experts: Do all roads lead to Rome? *Cognitive psychology* 32(1):49–96.

Mirijovsky, J. and S. Popelka. 2016. Evaluation of color settings in aerial images with the use of eye-tracking user study. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLI-B3:767–767. https://doi.org/10.5194/isprs-archives-XLI-B3-763-2016.

Montello, D. R. 2009. Cognitive research in GIScience: Recent achievements and future prospects. *Geography Compass* 3(5):1824–1840. https://doi.org/10.1111/j.1749-8198.2009.00273.x.

Montemayor, C. and H. H. Haladjian. 2017. Perception and cognition are largely independent, but still affect each other in systematic ways: Arguments from evolution and the consciousness-attention dissociation. *Frontiers in Psychology* 8:40. https://doi.org/10.3389/fpsyg.2017.00040.

Mulder, N. J. 1985. Decision making and classification. *Photogrammetria* 40(2):95–116. https://doi.org/10.1016/0031-8663(85)90004-3.

Nakashima, R., K. Kobayashi, E. Maeda, T. Yoshikawa and K. Yokosawa. 2013. Visual search of experts in medical image reading: The effect of training, target prevalence, and expert knowledge. *Frontiers in Psychology* 4:166. https://doi.org/10.3389/fpsyg.2013.00166.

Nicolin, B. and R. Gabler. 1987. A knowledge-based system for the analysis of aerial images. *IEEE Transactions on Geoscience and Remote Sensing* GE-25(3):317–329. https://doi.org/10.1109/TGRS.1987.289803.

Noble, I. R. 1987. The role of expert systems in vegetation science. *Vegetatio* 69(1–3):115–121. https://doi.org/10.1007/BF00038692.

Pannasch, S., J. R. Helmert, B. C. Hansen, A. M. Larson and L. C. Loschky. 2014. Commonalities and differences in eye movement behavior when exploring aerial and terrestrial scenes. In *Cartography from Pole to Pole: Selected Contributions to the XXVIth International Conference of the ICA, Dresden, 2013*, held in Dresden, German, 25–30 August 2013. https://doi.org/10.1007/978-3-642-32618-9_30.

Pesapane, F., P. Tantrige, F. Patella, P. Biondetti, L. Nicosia, A. Ianniello, U. G. Rossi, G. Carrafiello and A. M. Ierardi. 2020. Myths and facts about artificial intelligence: Why machine- and deep-learning will not replace interventional radiologists. *Medical Oncology* 37(5):40.

Phipps, M. and S. Rowe. 2010. Seeing satellite data. *Public Understanding of Science* 19(3):311–321. https://doi.org/10.1177/0963662508098684.

Pope, P. A. 2012. *Benchmark Imagery Project Merit Review, Reports LA-UR-12-00617 and LA-UR-12-617*. Los Alamos, N.M.: Los Alamos National Laboratory.

Rabben, E. L., E. L. Chalmers Jr, E. Manley and J. Pickup J. 1960. Fundamentals of photo interpretation. In *Manual of Photographic Interpretation*, edited by R. N. Colwell, 99–168. Washington, DC: American Society of Photogrammetry and Remote Sensing.

Radoux, J. and P. Bogaert. 2017. Good practices for object-based accuracy assessment. *Remote Sensing* 9(7):646. https://doi.org/10.3390/rs9070646.

Rajbhandari, S., J. Aryal, J. Osborn, R. Musk and A. Lucieer. 2017. Benchmarking the applicability of ontology in geographic object-based image analysis. *ISPRS International Journal of Geo-Information* 6(12):386. https://doi.org/10.3390/ijgi6120386.

Raubal, M. 2018. Cognition and geographic information technologies. In *Handbook of Behavioral and Cognitive Geography*, edited by D. R. Montello, 197–210. Location: Edward Elgar Publishing. https://doi.org/10.4337/9781784717544.00019.

Reichstein, M., G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais and Prabhat. 2019. Deep learning and process understanding for data-driven Earth system science. *Nature* 566(7743):195–204. https://doi.org/10.1038/s41586-019-0912-1.

Sadacca, R. 1963. Human factors in image interpretation. *Photogrammetric Engineering* 29(6):978–988.

Saralioglu, E. and O. Gungor. 2019. Use of crowdsourcing in evaluating post-classification accuracy. *European Journal of Remote Sensing* 52(sup1):137–147. https://doi.org/10.1080/22797254.2018.1564887.

Sarma, L.C.S. and V.V.S. Sarma. 1994. A prototype expert system for interpretation of remote sensing image data. *Sadhana* 19(1):93–111. https://doi.org/10.1007/BF02760392.

Schatzley, B. L. and L. S. Karably. 1954. An introduction to photo interpretation problems and research. *Photogrammetric Engineering* 20(3):428–443.

See, L., A. Comber, C. Salk, S. Fritz, M. van der Velde, C. Perger, C. Schill, I. McCallum, F. Kraxner and M. Obersteiner. 2013. Comparing the quality of crowdsourced data contributed by expert and non-experts. *PLoS One* 8(7):e69958. https://doi.org/10.1371/journal.pone.0069958.

Šikl, R., H. Svatoăová, F. Dăchtărenko and T. Urbánek. 2019. Visual recognition memory for scenes in aerial photographs: Exploring the role of expertise. *Acta Psychologica* 197:23–31. https://doi.org/10.1016/j.actpsy.2019.04.019.

Smart, P. R. 2018. Human-extended machine cognition. *Cognitive Systems Research* 49:9–23. https://doi.org/10.1016/j.cogsys.2017.11.001.Sowden, S., D. Stea, M. Blades, C. Spencer and J. M. Blaut. 1996. Mapping abilities of four-year-old children in York, England. *Journal of Geography* 95(3):107–111. https://doi.org/10.1080/00221349608978701.

Summerson, C. H. 1954. A philosophy for photo interpreters. *Photogrammetric Engineering* 20(3):396–397.

Svatonova, H. 2015. Aerial and satellite images in crisis management: Use and visual interpretation. In *International Conference on Military Technologies (ICMT) 2015*, held in Brno, Czech Republic, 19–21 May 2015. https://doi.org/10.1109/MILTECHS.2015.7153705.

Svatonova, H. and M. Rybansky. 2014. Children observe the Digital Earth from above: How they read aerial and satellite images. In *8th International Symposium of the Digital Earth (ISDE8)*, held in Kuching, Malaysia, 26–29 August 2013. https://doi.org/10.1088/1755-1315/18/1/012071.

Thagard, P. 2019. Cognitive Science. In *The Stanford Encyclopedia of Philosophy,* Spring 2019 ed., edited by E. N. Zalta. <https://plato.stanford.edu/archives/spr2019/entries/cognitive-science/> Accessed DD Month YYYY.

Tuia, D., C. Persello and L. Bruzzone. 2016. Domain adaptation for the classification of remote sensing data: An overview of recent advances. *IEEE Geoscience and Remote Sensing Magazine* 4(2):41–57. https://doi.org/10.1109/MGRS.2016.2548504.

Van Coillie, F.M.B., S. Gardin, F. Anseel, W. Duyck, L.P.C. Verbeke and R. R. De Wulf. 2014. Variability of operator performance in remote-sensing image interpretation: The importance of human and external factors. *International Journal of Remote Sensing* 35(2):754–778. https://doi.org/10.1080/01431161.2013.873152.

Verstappen, H. T. 1962–1964. Commission VII—photo interpretation report 1960–1964. *Photogrammetria* 19:137–143. https://doi.org/10.1016/S0031-8663(62)80087-8.

Vink, A. P. 1964. *Some Thoughts on Photo-Interpretation*. Location: International Institute for Aerial Survey and Earth Studies (ITC).

VoPham, T., J. E. Hart, F. Laden and Y.-Y. Chiang. 2018. Emerging trends in geospatial artificial intelligence (geoAI): Potential applications for environmental epidemiology. *Environmental Health* 17(1):40. https://doi.org/10.1186/s12940-018-0386-x.

Wardlaw, J., J. Sprinks, R. Houghton, J.-P. Muller, P. Sidiropoulos, S. Bamford and S. Marsh. 2018. Comparing experts and novices in Martian surface feature change detection and identification. *International Journal of Applied Earth Observation and Geoinformation* 64:354–364. https://doi.org/10.1016/j.jag.2017.05.014.

Wen, W., J. Timmermans, Q. Chen and P. M. van Bodegom. 2021. A review of remote sensing challenges for food security with respect to salinity and drought threats. *Remote Sensing* 13(1):6. https://doi.org/10.3390/rs13010006.

White, R. A., A. Çöltekin and R. R. Hoffman. 2018. *Remote Sensing and Cognition: Human Factors in Image Interpretation*. Boca Raton, Fla.: CRC Press. https://doi.org/10.1201/9781351040464.

Yusoff, N. M. and S. S. Salim. 2020. Shared mental model processing in visualization technologies: A review of fundamental concepts and a guide to future research in human-computer interaction." In *Engineering Psychology and Cognitive Ergonomics: Mental Workload, Human Physiology, and Human Energy*, held in Copenhagen, Denmark, 19–24 July 2020. https://doi.org/10.1007/978-3-030-49044-7_20.

Zelinsky, G. and J. Schmidt. 2009. An effect of referential scene constraint on search implies scene segmentation. *Visual Cognition* 17(6-7):1004–1028. https://doi.org/10.1080/13506280902764315.

Multi-View Urban Scene Classification with a Complementary-Information Learning Model

Wanxuan Geng, Weixun Zhou, and Shuanggen Jin

Abstract

Traditional urban scene-classification approaches focus on images taken either by satellite or in aerial view. Although single-view images are able to achieve satisfactory results for scene classification in most situations, the complementary information provided by other image views is needed to further improve performance. Therefore, we present a complementary information-learning model (CILM) to perform multi-view scene classification of aerial and ground-level images. Specifically, the proposed CILM takes aerial and ground-level image pairs as input to learn view-specific features for later fusion to integrate the complementary information. To train CILM, a unified loss consisting of cross entropy and contrastive losses is exploited to force the network to be more robust. Once CILM is trained, the features of each view are extracted via the two proposed feature-extraction scenarios and then fused to train the support vector machine classifier for classification. The experimental results on two publicly available benchmark data sets demonstrate that CILM achieves remarkable performance, indicating that it is an effective model for learning complementary information and thus improving urban scene classification.

Introduction

With the rapid development of remote sensing technology, traditional pixel-level image analysis has been unable to meet the needs of high-level image-content interpretation due to increasing spatial resolution, and urban scene classification has therefore been a hot topic in the remote sensing field (Zhou *et al.* 2018). Scene classification is assigning a specific label to each image according to its content (Kang *et al.* 2020), providing relatively high-level interpretation of a remote sensing image compared with pixel- and object-based classification (Xia *et al.* 2017). It is a practical application of high-resolution remote sensing image processing, which can provide data support for land planning and utilization (K. Xu *et al.* forthcoming), and is widely used in urban functional zoning planning (Huang *et al.* 2018), natural-disaster monitoring (Attari *et al.* 2018), and object detection (Schilling *et al.* 2018). Though the literature has developed a large number of scene-classification approaches—including handcrafted methods and ones based on deep learning—which can achieve remarkable performance, there are still problems to be solved.

On one hand, a high-resolution remote sensing image has rich spatial information and a complex background, making it difficult to extract powerful features for scene classification (T. Tian *et al.* 2021), and accordingly results in worse performance. On the other hand, most of the existing scene-classification approaches focus on images taken from a single view, such as satellite or aerial, but it has been demonstrated that the complementary information provided by other views is able to further improve classification performance (Machado *et al.* 2021), as shown in Figure 1. It is notable that scene classification of an aerial image can benefit from the complementary information provided by a ground-level image, and vice versa. For instance, we cannot obtain the correct classification result of an airport unless both aerial and ground-view images are exploited. In recent work by Machado *et al.* (2021), early and late fusion based on a convolutional neural network (CNN) are exploited to perform multi-view scene classification. More specifically, the early fusion is conducted by fusing the convolutional features of each view via a concatenation layer, whereas the late fusion is conducted by combining the prediction result of each view achieved by an individual CNN. Both early and late fusion have been proven effective for scene classification, but for early fusion, the concatenation layer is inserted in the first several convolutional layers, which cannot integrate the high-level features of each view image. For late fusion, an individual CNN must be trained for the prediction of each view image, and the training process is time-consuming and totally separated. We therefore raise the question: *Is it possible to learn complementary information via feature-level fusion and perform multi-view classification using a single CNN framework*?

Inspired by cross-view geo-localization (Vo and Hays 2016; T. Tian *et al.* 2021), in this article we extend our previous work (Geng *et al.* 2021) and propose a complementary information-learning model (CILM) for multi-view urban scene classification of aerial and ground-level images. The proposed CILM is a two-branch network trained using a unified loss to enhance the performance. Once CILM is trained, the high-level features of each view image are extracted and then combined to train a support vector machine (SVM) classifier to perform the final prediction. It should be noted that our work is different from that of Machado *et al.* (2021) in that, although both approaches take aerial and ground-level image pairs as input, for Machado *et al.* aerial and ground-level images in each pair are from the same location and the same class, whereas we ignore the location and the class of image pairs. Therefore, we explored how the information provided by pairs of images from different locations can benefit urban scene classification. Also, in our work, CILM is regarded as a feature extractor for extracting high-level features of each view image, which is not exploited for prediction. And we train an SVM classifier

Wanxuan Geng and Weixun Zhou are with the School of Remote Sensing and Geomatics Engineering, Nanjing University of Information Science and Technology, Nanjing, China (zhouwx@nuist.edu.cn).

Shuanggen Jin is with the School of Remote Sensing and Geomatics Engineering, Nanjing University of Information Science and Technology, Nanjing, China; and the China and Shanghai Astronomical Observatory, Chinese Academy of Sciences, Shanghai, China.

Contributed by Alper Yilmaz, August 30, 2021 (sent for review August 30, 2021.

Photogrammetric Engineering & Remote Sensing
Vol. 88, No. 1, January 2022, pp. 65–72.
0099-1112/22/65–72

doi: 10.14358/PERS.21-00062R2

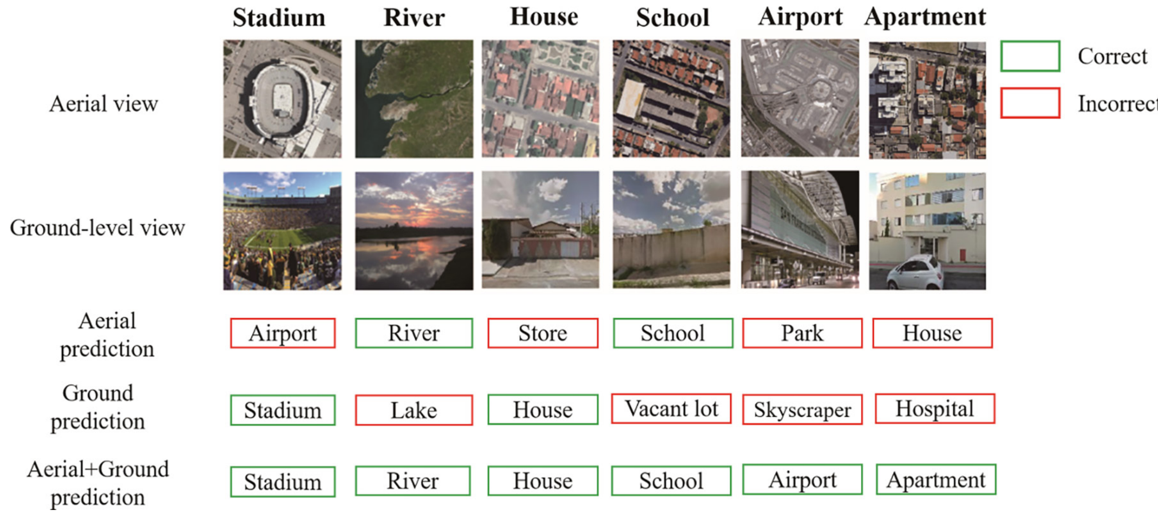


Figure 1. Visual examples of single- and multi-view scene classification.

using the fused high-level features to integrate complementary information for classification, which has been demonstrated to outperform the softmax classifier for scene classification (Xia *et al.* 2017).

In summary, the main contributions of this article are as follows.

- We propose a complementary information-learning model trained with a unified loss to integrate complementary information for multi-view scene classification of aerial and ground-level images. The unified loss is composed of cross entropy and contrastive losses, where the cross-entropy loss is to distinguish the class of each view image in the pair and to identify whether the input is a matched pair (i.e., aerial and ground-level images belonging to the same class) and the contrastive loss is to pull matched pairs closer and push unmatched pairs away in the feature space.
- We explore two pretrained CNNs as the basic network to construct CILM, which is then evaluated on two publicly available benchmark data sets with various experimental configurations, thus providing baseline results for future research.

The remainder of this paper is organized as follows. The next section reviews related work on urban scene classification. The proposed CILM is introduced in detail in the section after that, and then the experimental setup and results presented. Finally, we give a brief conclusion.

Related Work

In this section, we briefly review the work on scene classification and cross-modal methods for the processing of multi-view images.

Scene Classification

Traditional remote sensing scene classification is based on handcrafted low- and middle-level features. The low-level features are either global features, such as the color histogram (Swain and Ballard 1991), texture features (Haralick *et al.* 1973), and gist (Oliva and Torralba 2001), or local features, such as the famous scale-invariant feature transformation (Lowe 2004). In contrast, middle-level features establish the relationship with semantics through statistical-distribution analysis of low-level features; bag of visual words (Mansoori *et al.* 2013) is one of the representative methods, commonly used for classification tasks (Okumura *et al.* 2011). In recent years, methods based on deep learning have been widely exploited for scene classification, since CNNs outperform their counterpart traditional approaches on ImageNet (Krizhevsky *et al.* 2012), and have become the most popular approaches for image recognition since then. Zhou *et al.* (2017) proposed using a three-layer perceptron and a couple of convolutional layers to construct a low-dimensional CNN for remote sensing image retrieval. Han *et al.* (2017) integrated the pretrained AlexNet with spatial pyramid pooling and side supervision to improve scene-classification performance. Bian *et al.* (2017) proposed a simple yet effective saliency-patch sampling method to extract image regions that are the most informative.

Since effective and discriminative feature representation plays an important role in classification results (Zhang *et al.* 2019), some works focus on how to extract powerful features. Liu *et al.* (2018) rearranged deep features and used discriminative convolution filters with different kernel sizes for scene classification. Xu *et al.* (2020) used the transferred VGG16 to extract the multi-layer convolutional features and added several layers to process hierarchical features in different branches, which can improve performance; whereas Liu *et al.* (2018) combined spatial pyramid pooling with deep CNNs and designed a multiple-kernel learning strategy to fuse multi-scale features.

Though these handcrafted and particularly CNN feature-based methods have achieved significant success for scene classification, their data sources are single-view satellite or aerial images; whether the complementary information provided by other view images can benefit scene classification has not been explored.

Cross-Modal Approaches for Multi-View Images

A cross-modal network, as its name implies, is trained using more than one kind of data, and is a commonly used approach to process images of different views simultaneously. In work by X. Xu *et al.* (2015), the earliest cross-modal network was presented for image and text retrieval, which supports searching across multi-modal data and thus is suitable for remote sensing data (X. Xu *et al.* 2017). T. Tian *et al.* (2021) proposed an effective framework of cross-view matching for geo-localization in urban environments. Khokhlova *et al.* (2020) introduced a multi-modal network across time that learns to retrieve by content vertical aerial images of French urban and rural territories taken about 15 years apart. Xiong *et al.* (2020) proposed a novel deep cross-modality hashing network for cross-modal content-based remote sensing image retrieval between synthetic aperture radar and optical sensors. Feng *et al.* (forthcoming) proposed a framework for multi-view spectral–spatial feature extraction and fusion for analysis and classification of hyperspectral images. Xu *et al.* (2020) used hand-drawn sketches describing mental pictures to retrieve the desired targets in large-scale remote sensing images.

Differentiating our work here, most of the existing cross-modal works are essentially image matching to determine whether the input pairs are matched, such as the problem of image retrieval and geo-localization. The function of CILM, on the other hand, is to integrate the complementary information provided by each view image and then perform scene classification of multi-view images, which is a more difficult task than image matching.

Methodology

This section presents our methodology. We first introduce the architecture of the proposed CILM, then describe the unified loss used to train the network.

The Architecture of CILM

Our CILM consists of two identical subnetworks and three additional fully connected (FC) layers, as shown in Figure 2. The subnetwork is a CNN pretrained on ImageNet and contains convolution, pooling, and FC layers. CILM takes positive and negative image pairs as input, where a positive image pair is assigned the label 1 and a negative image pair is assigned the label 0. For positive image pairs, the aerial and ground-level images are from the same class, whereas for negative image pairs, they are from different classes.

During training, the aerial and ground-level images in a pair are each fed into one of the two subnetworks. The output feature vectors from each subnetwork are combined through a subtraction operation and the result is passed through the additional FC layer FC_{ag}, with a single output. We use a sigmoid function to convert this output value to a probability between 0 and 1, indicating the prediction of whether the input pairs are matched or unmatched. The first loss L1 is used for this task during training.

Relating to the other two additional FC layers, both FC_a and FC_g convert the 4096-D feature vectors from the subnetworks to N-D feature vectors, where N is the number of scene categories. Therefore, FC_a is used for aerial scene classification, whereas FC_g is used for ground-level scene classification. The motivation here is to force CILM to be more robust by using single-view image classification, which has been proven effective for scene classification (X. Liu *et al.* 2019). The second loss L_{2a} and L_{2g} are used for aerial and ground-level classification, respectively, during training.

The discriminative feature representation is significant for scene classification (Cheng *et al.* 2018); we therefore use the third loss L3 to learn powerful features. This is a ranking loss that can pull matched pairs closer and push unmatched pairs away in the feature space.

Once CILM is trained, we propose two scenarios to extract feature vectors to train the SVM classifier for classification, since SVM has been demonstrated to be more effective than the softmax classifier. More specifically, for the first scenario we extract features (i.e., f_a and f_g) from the last FC layers of the subnetworks, whereas for the second scenario we extract features (i.e., f_a' and f_g') from FC_a and FC_g. The extracted features are then fused to a feature vector through an addition operation.

Loss for CILM

The unified loss is exploited to update CILM during training. The unified loss L_U is defined as

$$L_U = \lambda_1 L_1 + \lambda_2 L_2 + \lambda_3 L_3 \tag{1}$$

where λ_1, λ_2, and λ_3 are three trade-off parameters that control the importance of these three losses.

L_1 is a binary cross-entropy loss defined as

$$L_1 = -q\log(p) - (1-q)\log(1-p) \tag{2}$$

$$p = \mathrm{sigmoid}(f_{ag}) \tag{3}$$

where q and p are the ground truth and the predicted label of the input pair, respectively, and f_{ag} is the output value of the FC_{ag} layer.

L_2 is a softmax cross-entropy loss consisting of two parts, L_{2a} for aerial-view classification and L_{2g} for ground-level view classification:

$$L_2 = L_{2a} + L_{2g} \tag{4}$$

$$L_{2a} = -\sum_{i=1}^{N} q_i^{a} \log\left(p_i^{a}\right) \tag{5}$$

$$L_{2g} = -\sum_{i=1}^{N} q_i^{g} \log\left(p_i^{g}\right) \tag{6}$$

where N is the number of scene categories, q_i^a and p_i^a are the ground truth and predicted label of the aerial image, and q_i^g and p_i^g are the ground truth and predicted label of the ground-level image.

L_3 is a contrastive loss aiming to compare the similarity between aerial and ground-level images in the pairs:

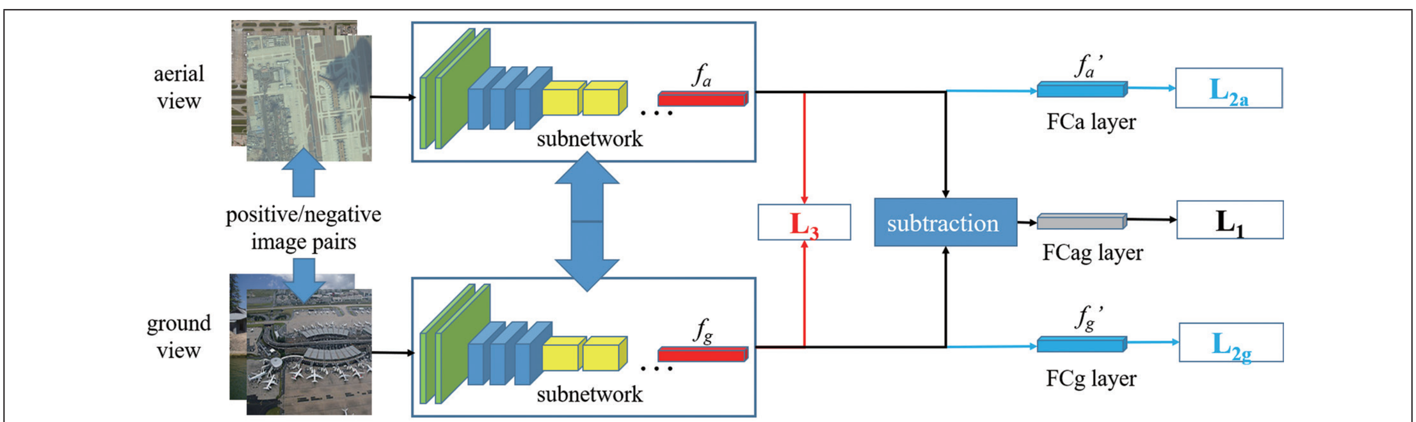


Figure 2. The network architecture of the proposed complementary information-learning model (CILM).

$$L_3 = \frac{1}{2} y d^2 + (1-y)\max(m-d,0)^2 \tag{7}$$

$$\mathrm{d} = \|f_a - f_g\|_2 \tag{8}$$

where y is the label of the input pair, d is the Euclidean distance between f_a and f_g, and m is the margin parameter used for constraint. If aerial and ground-level images in a pair are similar (i.e., the two images are from the same class), then d should be smaller than m; otherwise it is larger.

Experiments

In this section, we first describe two publicly available benchmark multi-view data sets, and then we introduce the experimental settings for our experiments. Finally, the experimental results and discussions are given.

Multi-View Data Sets

Our approach is evaluated using two benchmark data sets presented by Machado *et al.* (2021). The first, AiRound, is composed of 11 classes: airport, bridge, church, forest, lake, river, skyscraper, stadium, statue, tower, and urban park (Figure 3). Each class contains images in three distinct perspectives: satellite view, aerial view, and ground-level view. Therefore, each image in AiRound is composed of a triplet, with all three images acquired from the same place. Figure 4 shows some examples of image pairs; in our experiments, we use only the aerial and ground-level view images.

The second data set, CV-BrCT, is composed of approximately 24 000 pairs of images split into nine urban classes: apartment, hospital, house, industrial, parking lot, religious, school, store, and vacant lot (Figure 5). Each class has images in two distinct perspectives: aerial view and ground-level view. The two view images in each pair are also acquired from the same place. Figure 6 shows some examples of image pairs.

Experimental Setting

As described before, we did not consider whether the aerial and ground-level images in each pair were from the same location or the same class. In our experiments, we construct image pairs by first randomly splitting the images in each

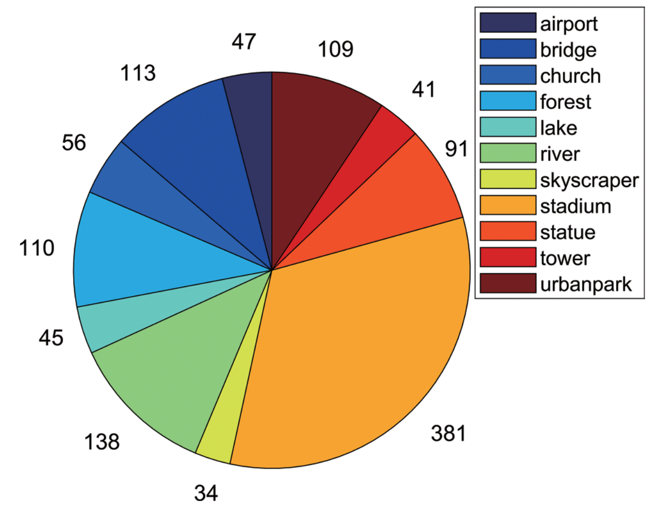


Figure 3. The number of images of each class in the AiRound data set.

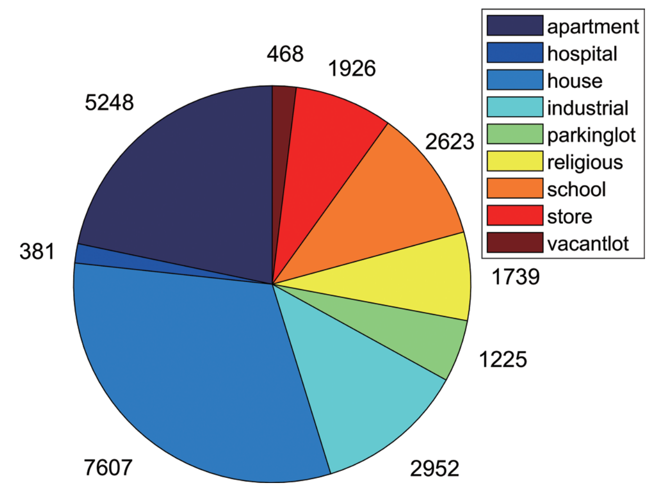


Figure 4. The number of images of each class in the CV-BrCT data set.

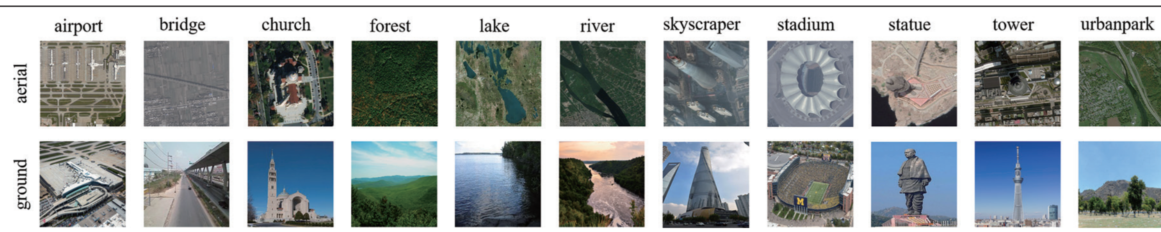


Figure 5. Examples of image pairs taken from the AiRound data set.

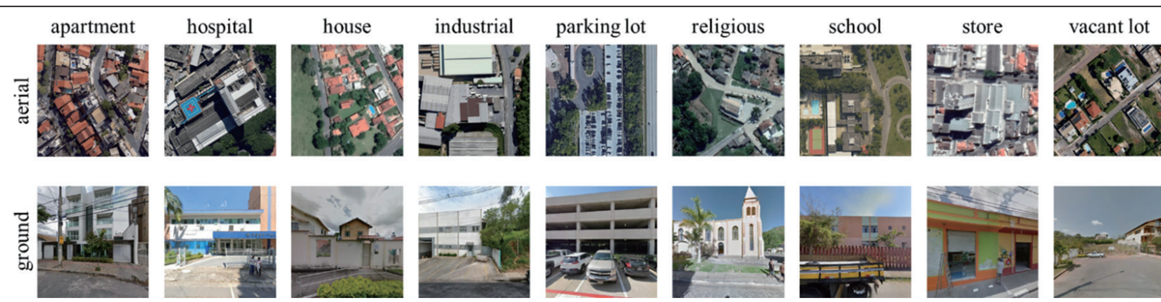


Figure 6. Examples of image pairs taken from the CV-BrCT data set.

Table 1. The training parameters of CILM on two data sets.

Data Set	Basic Network	Batch Size	Learning Rate	Iteration
AiRound	AlexNet	80	0.000 08	1000
	VGG16	24	0.000 08	1500
CV-BrCT	AlexNet	80	0.000 08	3000
	VGG16	24	0.000 08	5000

CILM = complementary information-learning model.

Table 2. The implementation details of single-view classification approaches.

Method	Implementation Details
CILM_1_2	CILM + L_1 and L_2 losses + subnetwork + softmax classifier
CILM_U	CILM + unified loss + subnetwork + softmax classifier

CILM = complementary information-learning model.

Table 3. The implementation details of multi-view classification approaches.

Method		Implementation Details
CILM_1	FS	CILM+L_1 loss + FS
CILM_1_3	FS	CILM + L_1 and L_3 losses + FS
CILM_1_2	FS	CILM + L_1 and L_2 losses + FS
	SS	CILM + L_1 and L_2 losses + SS
CILM_U	FS	CILM + unified loss + FS
	SS	CILM + unified loss + SS

class as 80% training samples and 20% test samples. Then we group aerial and ground-level images in each class through the method of exhaustion to obtain image pairs.

Regarding CILM, we select AlexNet (Krizhevsky *et al.* 2012) and VGG16 (Simonyan and Zisserman 2015) as the subnetworks, which are famous shallow and deep CNNs, respectively, that have been widely used for image classification. We remove the last FC layers in each to for the subnetworks to output 4096-D feature vectors. During training, the image pairs are resized to 227×227 pixels for AlexNet and 224×224 pixels for VGG16. The Adam optimizer is exploited to minimize the unified loss, where the gradient decay and the squared gradient decay factor are set to 0.9 and 0.99, respectively. The training details of CILM, such as batch size, learning rate, and number of iterations, are shown in Table 1. For the unified loss, we set $\lambda_1 = 1$, $\lambda_2 = 0.5$, $\lambda_3 = 0.0001$, and $m = 0.3$.

In the following experiments, we conduct single- and multi-view classification to evaluate the performance of CILM using the AiRound and CV-BrCT data sets. The single-view classification is aerial or ground-level classification using the subnetworks in CILM and the pretrained CNNs. Specifically, we evaluate the performance achieved by two CILM-based methods CILM_1_2 and CILM_U. The implementation details are shown in Table 2. Regarding the multi-view classification, we explore CILM with different configurations shown in Table 3.

In addition, CILM is compared to feature fusion and six-channel methods. Unless particularly stated, we extract features from the penultimate FC layer of the pretrained CNN and use SVM for classification.

Results on AiRound and CV-BrCT

Single-View Classification Results

The results of single-view classification obtained by CILM are presented to explore how the complementary information provided by other view images can benefit scene classification. All the results obtained by CILM are presented in Table 4.

For both the AiRound and CV-BrCT data sets, we can see that CILM_U configured with VGG16 (not shared weights) as the subnetworks achieves the best performance for both aerial and ground-level images. In addition, CILM trained other than with shared weights achieves slightly better performance than with shared weights, and VGG16 is a better subnetwork than AlexNet.

Multi-View Classification Results

Table 4. Single-view classification results of CILM on two data sets.

Weights	Subnetworks	Method	AiRound Aerial	AiRound Ground-level	CV-BrCT Aerial	CV-BrCT Ground-level
Shared	AlexNet	CILM_1_2	82.15	80.52	78.04	61.81
		CILM_U	82.83	81.82	78.39	62.39
	VGG16	CILM_1_2	83.69	81.97	79.46	62.64
		CILM_U	84.98	82.40	79.52	63.30
Not shared	AlexNet	CILM_1_2	84.55	82.40	78.09	63.16
		CILM_U	84.78	82.83	79.66	63.50
	VGG16	CILM_1_2	84.12	82.83	79.91	63.61
		CILM_U	**85.83**	**83.26**	**80.37**	**63.72**

CILM = complementary information-learning model.

Table 5. Multi-view classification results of CILM on the AiRound data set.

Weights	Subnetworks	CILM_1 FS	CILM_1 SS	CILM_1_3 FS	CILM_1_3 SS	CILM_1_2 FS	CILM_1_2 SS	CILM_U FS	CILM_U SS
Shared	AlexNet	87.55	—	87.98	—	90.56	90.99	91.42	92.27
	VGG16	88.41	—	88.84	—	91.20	91.55	91.85	92.70
Not shared	AlexNet	89.27	—	89.70	—	91.38	91.70	92.06	93.13
	VGG16	**89.70**	—	**90.10**	—	**90.92**	**91.83**	**92.49**	**93.56**

CILM = complementary information-learning model; FS = first scenario; SS = second scenario.

Table 6. Multi-view classification results of CILM on the CV-BrCT data set.

Weights	Subnetworks	CILM_1 FS	CILM_1 SS	CILM_1_3 FS	CILM_1_3 SS	CILM_1_2 FS	CILM_1_2 SS	CILM_U FS	CILM_U SS
Shared	AlexNet	80.24	—	80.50	—	80.70	81.62	81.58	82.18
	VGG16	81.80	—	81.90	—	83.62	84.06	83.97	84.24
Not shared	AlexNet	80.52	—	80.60	—	81.66	82.11	82.38	82.42
	VGG16	**82.35**	—	**82.45**	—	**83.66**	**84.09**	**84.15**	**84.32**

CILM = complementary information-learning model; FS = first scenario; SS = second scenario.

Here we present the results of multi-view classification on the AiRound (Table 5) and CV-BrCT (Table 6) data sets obtained by the proposed CILM with different configurations. It can be observed that CILM_U configured with VGG16 (not shared weights) as the subnetworks and SS as the feature-extraction strategy achieves the best performance for both data sets. The results will be analyzed in detail.

It can be seen that CILMs not trained with shared weights achieve slightly better performance than those with shared weights, except for CILM_1_2 configured with VGG16 and the first scenario for the AiRound data set. The results make sense, since aerial and ground-level images are taken from different perspectives, and thus we can learn view-specific features when the subnetworks do not use shared weights. For the subnetworks, it seems that VGG16 is a better choice than AlexNet, but the performance difference is small. To explore how the proposed unified loss can improve the performance of CILM, we trained CILM using different losses. It is obvious that CILM_U outperforms the other approaches, indicating that the unified loss can benefit multi-view classification. We can also conclude that L2 is the most important among the three losses, according to the results obtained by CILM_1_2, CILM_1_3, and CILM_1. In addition, SS is a more appropriate feature-extraction scenario for CILM. This is because the first scenario extracts 4096-D features from the last FC layers of the subnetworks, whereas the second scenario extracts *N*-D features from the additional FC layers, where the features are class-specific high-level features, thus achieving better performance.

According to the results of multi- and single-view classification, we can conclude that multi-view scene classification can benefit from the complementary information provided by aerial or ground-level images. For AiRound, the best performance is 93.56, whereas the best single-view performance is 85.83 for the aerial view and 83.26 for the ground-level view. With respect to CV-BrCT, the best performance is 84.32, whereas the best single-view performance is 80.37 for the aerial view and 63.72 for the ground-level view. Therefore, multi-view classification improves the results of single-view classification by a significant margin, especially for the ground-level classification of CV-BrCT. This is possibly because the ground-level images in CV-BrCT are more challenging than the aerial images, as shown in Figure 6.

Feature-Visualization Results

In addition to the single- and multi-view classification results, we also present the visualization results of features extracted by CILM to give a quantitative evaluation, as can be observed in Figures 7 and 8. For the AiRound data set, the features of multi-view images can be easily separated for different classes, whereas for single-view images, most of the image classes are clustered together—except for stadium. Regarding the CV-BrCT data set, we can observe similar results as with AiRound. But an interesting phenomenon is that the features of multi-view images and aerial images achieve similar clustering performance, both outperforming ground-level images by a significant margin. These results make sense, since

Table 7. Performance comparisons of CILM and counterpart approaches for single- and multi-view classification.

	Single-View Classification			
	AiRound		CV-BrCT	
Method	**Aerial**	**Ground**	**Aerial**	**Ground**
CNN-softmax (Simonyan and Zisserman 2015)	82.84	81.55	79.18	62.12
CNN-SVM (Simonyan and Zisserman 2015)	80.52	80.09	69.87	54.95
CILM	**85.83**	**83.26**	**80.37**	**63.72**
	Multi-View Classification			
Method	**AiRound**		**CV-BrCT**	
Feature fusion (Simonyan and Zisserman 2015)	90.4		74.99	
Six-channel (Vo and Hays 2016)	70.39		73.46	
CILM	**93.56**		**84.32**	

CILM = complementary information-learning model; CNN = convolutional neural network.

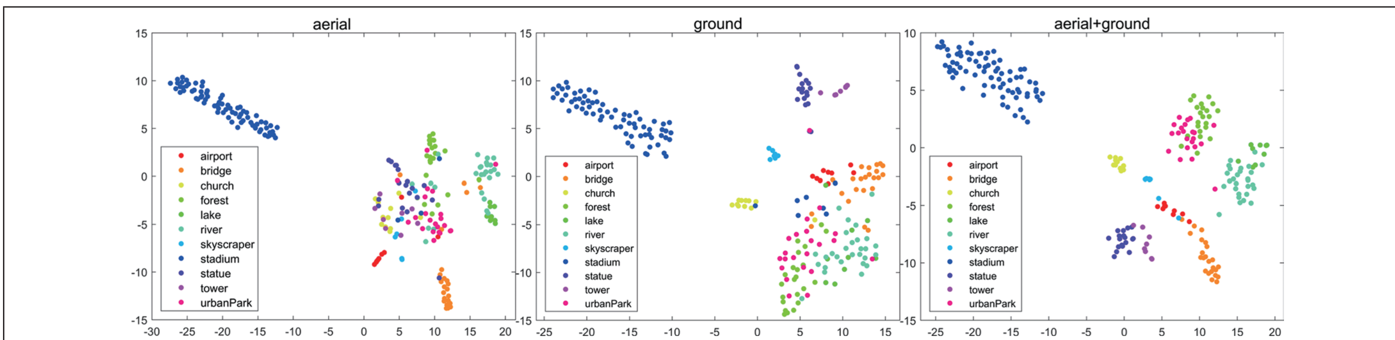


Figure 7. Feature-visualization results of single- and multi-view images in the AiRound data set.

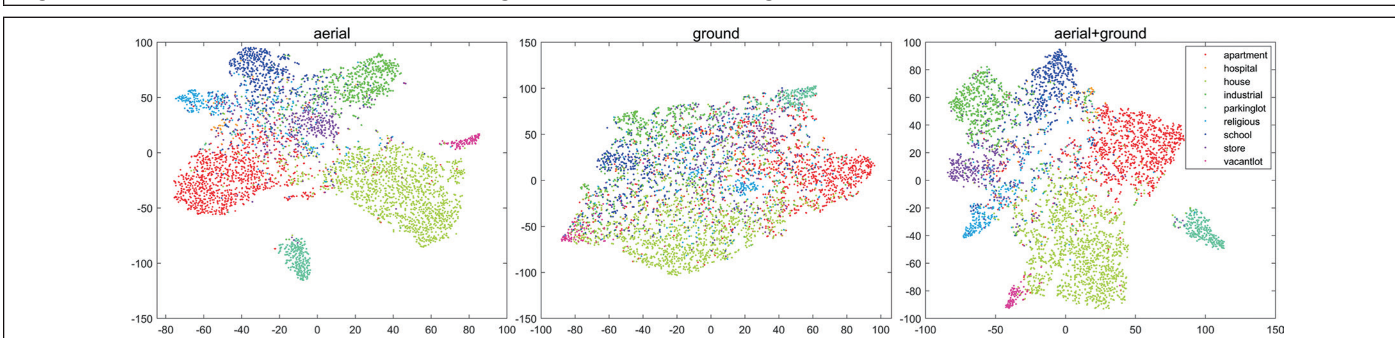


Figure 8. Feature-visualization results of single- and multi-view images in the CV-BrCT data set.

CV-BrCT is more challenging than AiRound, and the ground-level images in CV-BrCT have higher intraclass diversity.

Table 7 shows the comparison results of single- and multi-view classification achieved by CILM and other counterpart approaches on AiRound and CV-BrCT. For the multi-view classification, our method outperforms feature fusion and six-channel methods for both data sets. The six-channel method performs the worst among these approaches; is not as effective, as it was used for geo-localization (Vo and Hays 2016). This is because for image geo-localization, we only need to determine whether the two images are from the same location, whereas for multi-view classification we need to identify the classes of image pairs, which is definitely a more challenging problem. As for the single-view classification, our approach achieves better performance than the two pretrained CNN-based approaches for both data sets.

The confusion matrices of the multi-view results achieved by our approach on AiRound and CV-BrCT are shown in Figures 9 and 10, respectively. For AiRound, the classification accuracy of lake is below 0.8, and around 22% of lake samples are incorrectly classified to rivers due to the high similarity and the imbalanced number of samples between lake and river. Skyscraper also has a lower classification accuracy, due to the small number of samples, and some images are mistakenly classified in other building categories, such as airport and stadium. In addition, urban park is easily confused with forest. For CV-BrCT, the high similarity between different classes and the number of samples has a great influence on the classification accuracy. We can see that the classification accuracy of hospital is only 18%, because hospital is severely confused with apartment.

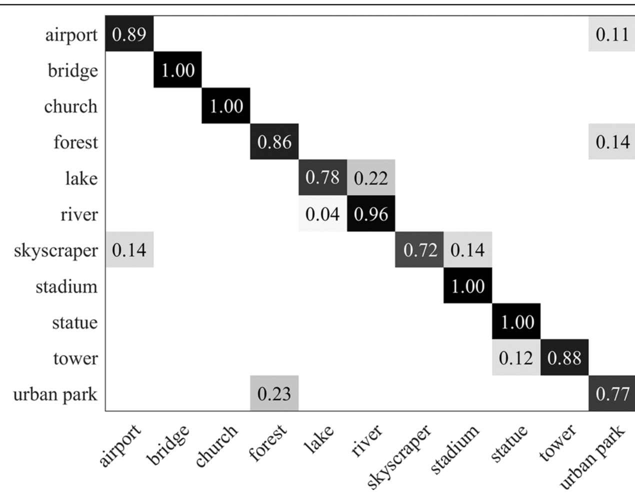


Figure 9. The confusion matrix for the multi-view classification results achieved by CILM on AiRound.

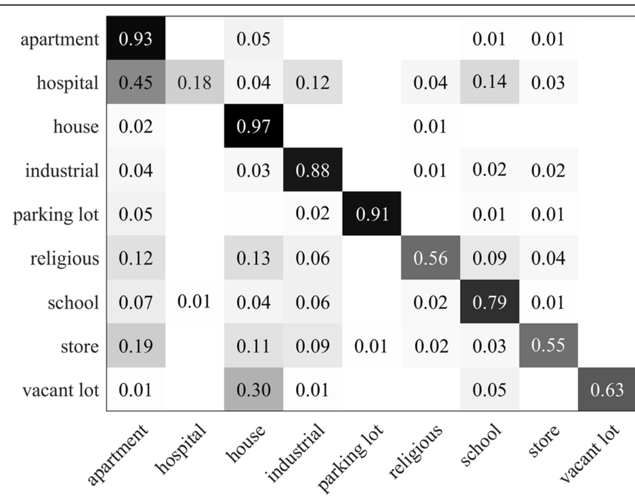


Figure 10. The confusion matrix for the multi-view classification results achieved by CILM on CV-BrCT.

Conclusion

In this article, we proposed a complementary information-learning model (CILM) for multi-view urban scene classification. To enhance the training of CILM, we exploited a unified loss consisting of two cross-entropy losses and a contrastive loss. Unlike the existing works that use softmax for classification, we extract the high-level features of aerial and ground-level images via two feature-extraction scenarios, and then fuse the features to integrate complementary information to train an SVM for classification. We explored CILM with different configurations of subnetworks, losses, and feature-extraction scenarios to evaluate its performance. The experimental results show that CILM configured with VGG16 (weights not shared) as the subnetworks and the second scenario as the feature-extraction strategy achieves the best performance on both AiRound and CV-BrCT data sets. Further, the comparison results between multi- and single-view classification indicate that the complementary information provided by other view images can benefit scene classification.

Acknowledgments

This work was supported by the Strategic Priority Research Program Project of the Chinese Academy of Sciences under grant XDA23040100; the National Natural Science Foundation of China under grant 42001285; the Natural Science Foundation of Jiangsu Province, China, under grant BK20200813; the Natural Science Foundation of the Jiangsu Higher Education Institutions of China under grant 20KJB420002; and the Jiangsu Double Innovation Doctor Program under grant R2020SCB58.

The authors would like to thank the anonymous reviewers for their comments to improve the article, and to thank the researchers whose work presents the AiRound and CV-BrCT data sets.

References

Attari, N., F. Ofli, M. Awad, J. Lucas and S. Chawla. 2017. Nazr-CNN: Fine-grained classification of UAV imagery for damage assessment. Pages 50–59 in *2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, held in Tokyo, Japan, 19–21 October 2017. Piscataway, NJ: IEEE.

Bian, X., C. Chen, L. Tian and Q. Du. 2017. Fusing local and global features for high-resolution scene classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10(6):2889–2901.

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.

Feng, J., J. Zhang and Y. Zhang. Forthcoming. A multiview spectral-spatial feature extraction and fusion framework for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*.

Geng. W, W. Zhou and S. Jin. 2021. Feature fusion for cross-modal scene classification of remote scene image. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLIV-M-3-2021:63–66. https://doi.org/10.5194/isprs-archives-XLIV-M-3-2021-63-2021.

Han, X., Y. Zhong, L. Cao and L. Zhang. 2017. Pre-trained AlexNet architecture with pyramid pooling and supervision for high spatial resolution remote sensing image scene classification. *Remote Sensing* 9(8):848.

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.

Huang, B., B. Zhao and Y. Song. 2018. Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery. *Remote Sensing of Environment* 214:73–86.

Kang, J., R. Fernandez-Beltran, Z. Ye, X. Tong, P. Ghamisi and A. Plaza. 2020. Deep metric learning based on scalable neighborhood components for remote sensing scene characterization. *IEEE Transactions on Geoscience and Remote Sensing* 58(12):8905–8918.

Khokhlova, M., V. Gouet-Brunet, N. Abadie and L. Chen. 2020. Cross-year multi-modal image retrieval using Siamese networks. Pages 1–5 in *2020 IEEE International Conference on Image Processing (ICIP)*, held in Abu Dhabi, United Arab Emirates, 25–28 October 2020. Piscataway, NJ: IEEE.

Krizhevsky, A., I. Sutskever and G. E. Hinton. 2012. ImageNet classification with deep convolutional neural networks. *Advances in neural information processing systems.* pp. 1097-1105.

Liu, N., X. Lu, L. Wan, H. Huo and T. Fang. 2018. Improving the separability of deep features with discriminative convolution filters for RSI classification. *ISPRS International Journal of Geo-Information* 7(3):95.

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, X., Y. Zhou, J. Zhao, R. Yao, B. Liu and Y. Zheng. 2019. Siamese convolutional neural networks for remote sensing scene classification. *IEEE Geoscience and Remote Sensing Letters* 16(8):1200–1204.

Lowe, D. G. 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60(2):91–110.

Machado, G., E. Ferreira, K. Nogueira, H. Oliveira, M. Brito, P. H. T. Gama and J. A. dos Santos. 2021. AiRound and CV-BrCT: Novel multiview datasets for scene classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14:488–503.

Mansoori, N. S, M. Nejati, P. Razzaghi and S. Samavi. 2013. Bag of visual words approach for image retrieval using color information. Pages 1–6 in *2013 21st Iranian Conference on Electrical Engineering (ICEE)*, held in Mashhad, Iran, 14–16 May 2013. Piscataway, NJ: IEEE.

Okumura, S., N. Maeda, K. Nakata, K. Saito, Y. Fukumizu and H. Yamauchi. 2011. Visual categorization method with a Bag of PCA packed Keypoints. Pages 950–953 in *2011 4th International Congress on Image and Signal Processing*, held in Shanghai, China, 15–17 October 2013. Piscataway, NJ: IEEE.

Oliva, A. and A. Torralba. 2001. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision* 42(3):145–175.

Schilling, H., D. Bulatov, R. Niessner, W. Middelmann and U. Soergel. 2018. Detection of vehicles in multisensor data via multibranch convolutional neural networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 11(11):4299–4316.

Simonyan, K. and A. Zisserman. 2015. Very deep convolutional networks for large-scale image recognition. Pages 1–14 in *International Conference on Learning Representations (ICLR)*, held in San Diego, California, US, 7-9 May 2015. New York: Elsevier.

Swain, M. J. and D. H. Ballard. 1991. Color indexing. *International Journal of Computer Vision* 7(1):11–32.

Tian, T., L. Li, W. Chen and H. Zhou. 2021. SEMSDNet: A multiscale dense network with attention for remote sensing scene classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14:5501–5514.

Tian, Y., C. Chen and M. Shah. 2017. Cross-view image matching for geo-localization in urban environments. Pages 1998–2008 in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, held in Honolulu, HI, USA, 21-26 July 2017. Piscataway, NJ: IEEE.

van der Maaten, L. and G. Hinton. 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research* 9:2579–2605.

Vo, N. N. and J. Hays. 2016. Localizing and orienting street views using overhead imagery. Pages 494–509 in *Computer Vision–ECCV 2016*. Lecture Notes in Computer Science vol. 9905. Cham, Switzerland: Springer.

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.

Xiong, W., Z. Xiong, Y. Zhang, Y. Cui and X. Gu. 2020. A deep cross-modality hashing network for SAR and optical remote sensing images retrieval. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13:5284–5296.

Xu, F., W. Yang, T. Jiang, S. Lin, H. Luo and G.-S. Xia. 2020. Mental retrieval of remote sensing images via adversarial sketch-image feature learning. *IEEE Transactions on Geoscience and Remote Sensing* 58(11):7801–7814.

Xu, K., H. Huang, P. Deng and Y. Li. Forthcoming. Deep feature aggregation framework driven by graph convolutional network for scene classification in remote sensing. *IEEE Transactions on Neural Networks and Learning Systems*.

Xu, K., H. Huang, Y. Li and G. Shi. 2020. Multilayer feature fusion network for scene classification in remote sensing. *IEEE Geoscience and Remote Sensing Letters* 17(11):1894–1898.

Xu, X., F. Shen, Y. Yang, H. T. Shen and X. Li. 2017. Learning discriminative binary codes for large-scale cross-modal retrieval. *IEEE Transactions on Image Processing* 26(5):2494–2507.

Xu, X., Yang Y, Shimada A, Taniguchi R.I and He L. 2015. Semi-supervised coupled dictionary learning for cross-modal retrieval in Internet images and texts. Pages 847–850 in *MM' 15: Proceedings of the 23rd ACM International Conference on Multimedia*, held in Brisbane, Australia, 26–30 October 2015. Edited by Zhou, J. New York: Association for Computing Machinery.

Zhang, B., Y. Zhang and S. Wang. 2019. A lightweight and discriminative model for remote sensing scene classification with multidilation pooling module. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 12(8):2636–2653.

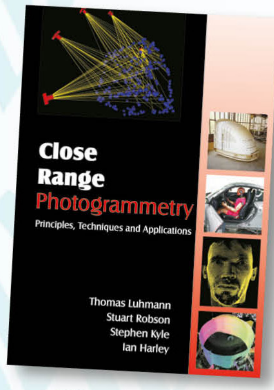
Zhou, W., S. Newsam, C. Li and Z. Shao. 2018. PatternNet: A benchmark dataset for performance evaluation of remote sensing image retrieval. *ISPRS Journal of Photogrammetry and Remote Sensing* 145(Part A):197–209.

Zhou, W., S. Newsam. C. Li and Z. Shao. 2017. Learning low dimensional convolutional neural networks for high-resolution remote sensing image retrieval. *Remote Sensing* 9(5):489.

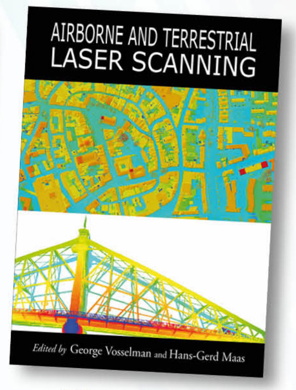
WHITTLES PUBLISHING | SIMPLY BRILLIANT BOOKS

ASPRS MEMBER EXCLUSIVE 20% DISCOUNT ON ALL WHITTLES PUBLISHING BOOKS USING CODE WPASPRS2

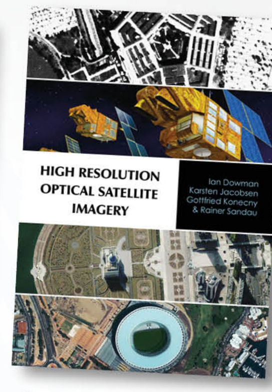
WHITTLES PUBLISHING'S STABLE OF CLASSIC GEOMATICS BOOKS INCLUDES THE THREE WINNERS OF THE PRESTIGIOUS KARL KRAUS MEDAL AWARDED BY ISPRS



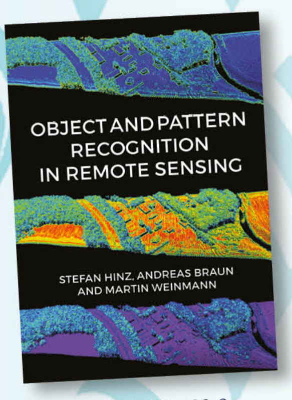
978-1870325-73-8
(Available as a CD)



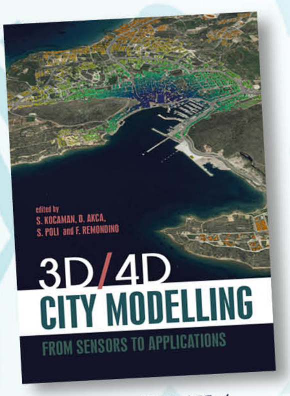
978-1904445-87-6



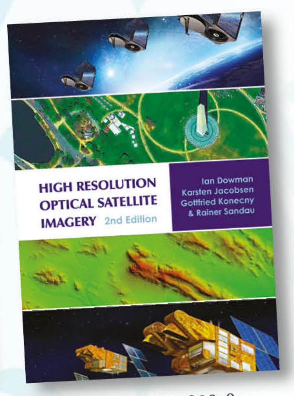
ISBN? Webite only has 2nd edition



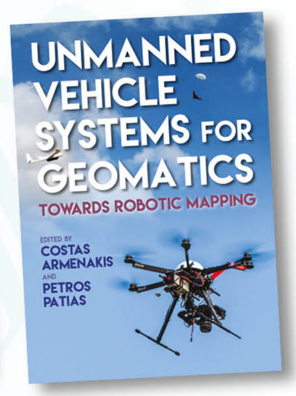
978-184995-128-9



978-184995-475-4



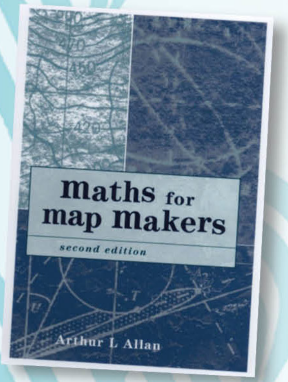
978-184995-390-0



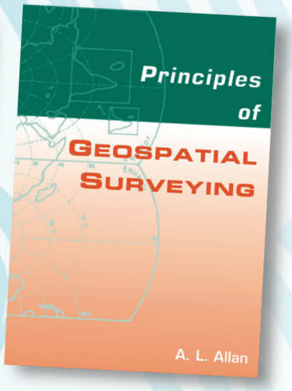
978-184995-127-2

OUR LIST CONTINUES TO EXPAND WITH THESE NEW TITLES

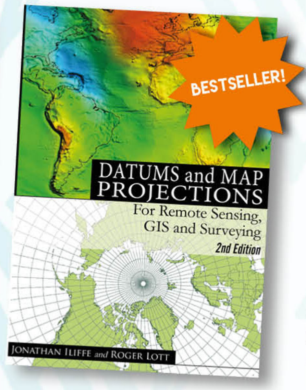
BROWSE OUR WEBSITE TO SEE OUR FULL RANGE OF ACCLAIMED GEOMATICS BOOKS



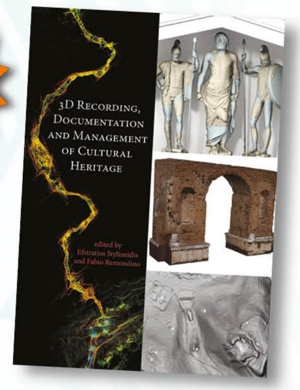
978-1870325-99-8



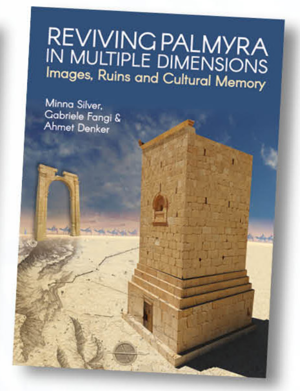
978-1904445-21-0



978-1904445-47-0



978-184995-168-5



978-184995-296-5

WWW.WHITTLESPUBLISHING.COM