XXIV ISPRS Congress
International Society for Photogrammetry and Remote Sensing
4 – 10 July 2021, Nice, France

www.isprs2020-nice.com

Keynote speakers

Executive committee

Platinum sponsor
AIRBUS

Gold sponsors
ThalesAlenia Space
esri

Silver sponsors
HEXAGON

Bronze sponsors

Institutional partners, supporting graduate schools and associations
ANNOUNCEMENTS

Avineon, Inc. (Avineon), a global provider of spatial intelligence, digital modernization, and engineering support services, is pleased to announce its recognition by Esri as a Cornerstone Partner. The announcement was made by Robert Laudati and Richard Cooke from Esri, during the closing session of the 2021 Esri Partner Conference.

The Cornerstone Partnership award recognizes organizations that have participated in the Esri business partner program for 20 years while demonstrating a commitment to improving the Esri and broader geospatial communities. Avineon has been working with Esri products since shortly after the company was founded in 1992. Avineon contributed to Esri’s expansive growth by assisting hundreds of utility and government customers in implementing Esri products. The company is also an Esri Gold Business Partner and recipient of the Esri Utility Network Specialty designation.

“Congratulations to Avineon on receiving the Esri Cornerstone Partner award at Esri Partner Conference 2021,” said Robert Laudati, Director, Global Partners and Alliances, Esri. “The company’s ability to combine extensive technical knowledge of Esri technology with understanding the needs of enterprise businesses, has enabled them to become a trusted Esri partner around the world. Avineon’s ongoing commitment to the partnership has truly been an asset in serving Esri users for more than 20 years.”

“This award reflects Avineon’s commitment to the spatial intelligence community at every level of the organization,” said Joel Campbell, Avineon’s Senior Vice President – Commercial Systems. “Our leadership has encouraged our employees to pursue Esri training and certifications, as well as funding participation in the partnership program and user conferences. Our consultants and employees have responded with enthusiasm in helping our customers get the most of their investments in Esri technology. We look forward to another 20 years of supporting the Esri community.”

For more information, visit www.avineon.com and esri.com.

CT Consultants, Inc. has purchased a Phase One 280MP Aerial Solution (PAS 280) 4-band, a large-format aerial imaging system, positioning the firm to expedite service with greater quality imagery to clients. This Phase One 280MP Aerial Solution is the first in the Midwest and adds CT to an elite number of companies owning this solution in the United States.

CT is a premier provider of advanced photogrammetric services, providing clients with unmatched technical quality, employing personnel with high academic and professional credentials, and continuously investing in new technologies. Their staff includes ASPRS-certified photogrammetrists, geographers, computer analysts, FAA-certified pilots, FAA-Part 107 drone pilots, aerial photographers, and geospatial technicians.

“Our geospatial team plays an integral part of our overall planning, design, and construction services. New technology investment is key as we provide clients with the necessary information and creative solutions,” says Dave Wiles, PE, CT’s President.

CT’s geospatial group uses unique data gathering and mapping tools, including aerial (film and digital) photography, aerial and mobile lidar systems, handheld lidar (SLAM), remotely piloted aircraft (drone), airborne GPS, and terrestrial lidar laser scanning while being anchored by boots on the ground land surveying. The firm, founded in 1922, has grown significantly in recent years through successful mergers and acquisitions, and they currently employ 237 professionals. Visit CT Consultants at cconsultants.com/services/geospatial to learn more.

For more information about Phase One, please visit geospatial.phaseone.com.

Extensis®, a leading provider of digital asset management solutions, today announced its GeoViewer Pro for Desktop application will now be offered at no charge. GeoViewer Pro is a standalone GIS image viewer for compressed MrSID files, raster imagery, lidar point clouds, and vector layers.

“GeoViewer Pro enables anyone – whether a GIS professional or map consumer – to view and interact with geospatial data sets without a full GIS software package,” said Terry Ryan, Federal Government Sales Manager at Extensis.

GeoViewer Pro was developed by LizardTech, now Extensis, as a paid multi-feature upgrade to the free GeoViewer application. It allows end users to view satellite, aerial and drone image files, as well as lidar data and GIS vectors. Moreover, it is a free way to view MrSID, JPEG 2000, NITF, LAS, and LAZ files created with the popular GeoExpress software.

“The GeoViewer application has been embraced by thousands of users worldwide, especially by GIS managers who want to fully leverage the value of their geospatial information by providing a no-cost way for users and the general public to view and interact with their image and lidar files,” said Ryan.

In addition to supporting a variety of raster image formats, GIS shapefiles, and even KML files, the core functionality of GeoViewer includes customizable display settings, support for OGC map services, exporting files, and displaying maps in their native projections. Users can make and save measurements and annotate geospatial files create new images or maps from the larger data sets for exporting and sharing with another user.

Current GeoViewer users will be able to upgrade to the new-free Pro version without a new download. First-time GeoViewer Pro users are invited to download the upgraded application to their desktop computers at extensis.com/geoviewer-pro.

Owned by the same parent company, Extensis and LizardTech united under the Extensis brand in September 2018, combining the industry-leading geospatial image management and compression expertise of LizardTech with the Digital Asset Management (DAM) capabilities of Extensis. These data management solutions enable clients to easily organize, access and share geospatial datasets along with associated documents.

For more information, visit www.extensis.com.
Dubai Municipality announced the successful launch and deployment of DMSat-1, an atmospheric monitoring microsatellite built by Space Flight Laboratory (SFL). DMSat-1 launched today from the Baikonur Cosmodrome in Kazakhstan aboard a Soyuz rocket.

SFL developed DMSat-1 under contract to the Dubai-based Mohammed Bin Rashid Space Centre (MBRSC) in the United Arab Emirates (UAE). The 15-kg microsatellite was built on SFL’s space-proven Next-generation Earth Monitoring and Observation (NEMO) platform.

“We congratulate Dubai on the launch of its first atmospheric monitoring nanosatellite,” said SFL Director, Dr. Robert E. Zee. “DMSat-1 will play a key role in monitoring atmospheric aerosols and greenhouse gases for Dubai and the UAE.

SFL was selected to build DMSat-1 for its compact size and performance, including the mission-critical importance of attitude control and precise sensor pointing. SFL has developed high-performance ground target tracking capabilities that enable the small satellite to execute a slewing maneuver in orbit to accurately point its sensors at selected swaths of the atmosphere.

DMSat-1 carries two instruments onboard. The primary payload is a multispectral polarimeter that monitors aerosols, which are typically fine particles of liquids and solids in the upper atmosphere often caused by anthropomorphic sources. Aerosols also correlate to natural phenomenon such as dust storms. The secondary payload is a pair of spectrometers for detection of greenhouse gases, such as carbon dioxide and methane.

DMSat-1 is the 16th SFL-built satellite launched in the past seven months. These launches include missions developed for GHGSat Inc. of Montreal, HawkEye 360 of Virginia, Space-SI of Slovenia, and a Toronto-based telecommunications company. Today’s launch also included two communications satellites developed using SFL technology.

For more information, visit www.utias-sfl.net.

TECHNOLOGY

Phase One today announced the P3 Payload, a versatile solution designed for fast, efficient, and safe inspection of critical infrastructure with an Unmanned Aerial Vehicle (UAV). The P3 Payload includes a Phase One iXM 100MP or 50MP camera, one of the RSM lens options, and a new gimbal with integrated rangefinder.

Available in Q2 2021, the P3 Payload is extremely versatile, offered in two configurations to meet a wide range of user requirements. The P3 for the DJI M300 is plug-and-play and ready to fly with a M300 drone, developed mainly for operators and service providers that already own, or are considering purchasing, a DJI M300. The P3 MAVLink is ready for integration with the many drones that are based on the open-source MAVLink protocol, ideal for bundling by OEMs and solution integrators.

The centerpiece of the payload is the Phase One iXM 100MP or 50MP camera. Robust and waterproof, these sensors boast a dynamic range that guarantees sharp image collection in high-contrast or low-light environments. The variety of lens options ensures large surface areas can be captured with millimeter-level detail – even at safe distances from the asset. The new gimbal with the integrated laser rangefinder ensures precise and fast focusing on every shot, eliminating blurry and out-of-focus images.

The Phase One P3 Payload is a turnkey solution to handle all your UAV inspection needs.

For more information, visit https://geospatial.phaseone.com/drone-payload/p3-payload-for-drones/.

UP42 will offer automated geospatial processing analytics from CATALYST on the UP42 marketplace and developer platform. Available in March, the first offering will be the CATALYST InSAR processing block, which measures millimeter-scale ground deformation from time series Synthetic Aperture Radar (SAR) satellite data sets.

Introduced in 2020, CATALYST is a new brand from PCI Geomatics. CATALYST launched the CATALYST Insights suite to package its algorithms and analytics tools into integrated, customizable processes. Designed for both geospatial professionals and non-technical users, these analytics extract actionable information from image data and present results in readily understandable formats.

“Our partnership with CATALYST will give UP42 users access to a range of powerful processing and analytics capabilities for satellite data,” said UP42 CEO Sean Wiid. “The initial InSAR processing block detects subtle terrain changes known as ground deformation. This has a range of vital applications related to monitoring critical infrastructure, tracking subsurface water levels, and predicting geohazards.”

The UP42 platform offers extensive Earth observation data sets and advanced processing algorithms – along with cloud computing power – to create custom geospatial solutions easily and inexpensively. Users purchase just the data needed to cover their area of interest and then leverage off-the-shelf processing capabilities to analyze the datasets without investment in their own computing infrastructure.

UP42 users will be able to apply the CATALYST InSAR analytics to Sentinel-1 satellite data available on the developer platform, as well as other SAR data sets in the future.

For land deformation detection, UP42 users will find the InSAR processing block is a valuable monitoring tool in protecting critical infrastructure by detecting small terrain changes before potentially catastrophic ground movement. InSAR measures subsidence – slumping and slope creep that pose risks to oil & gas wells, hydrocarbon pipelines, dams, mining sites, transportation corridors, and agriculture irrigation systems.

For more information, visit www.up42.com.

CALENDAR

• 7-11 June, URISA GIS Leadership Academy, Minneapolis, Minnesota. For more information, visit www.urisa.org/education-events/urisa-gis-leadership-academy/.
311 Inversion of Solar-Induced Chlorophyll Fluorescence Using Polarization Measurements of Vegetation
Haiyan Yao, Ziying Li, Yang Han, Haofang Niu, Tianyi Hao, and Yuyu Zhou

In vegetation remote sensing, the apparent radiation of the vegetation canopy is often combined with three components derived from different parts of vegetation that have different production mechanisms and optical properties: volume scattering $L_{vol}$, polarized light $L_{pol}$, and chlorophyll fluorescence $ChlF$. The chlorophyll fluorescence plays a very important role in vegetation remote sensing, and the polarization information in vegetation remote sensing has become an effective way to characterize the physical characteristics of vegetation. This article analyzes the difference between these three types of radiation flux and utilizes polarization radiation to separate them from the apparent radiation of the vegetation canopy.

339 Quality Assessment of Heterogeneous Training Data Sets for Classification of Urban Area with Landsat Imagery
Neema Nicodemus Lyimo, Fang Luo, Qimin Cheng, and Hao Peng

Quality assessment of training samples collected from heterogeneous sources has received little attention in the existing literature. Inspired by Euclidean spectral distance metrics, this article derives three quality measures for modeling uncertainty in spectral information of open-source heterogeneous training samples for classification with Landsat imagery.

349 Comparative Assessment of Target-Detection Algorithms for Urban Targets Using Hyperspectral Data
Shalini Gakhar and K.C. Tiwari

Hyperspectral data present better opportunities to exploit the treasure of spectral and spatial content that lies within their spectral bands. Hyperspectral data are increasingly being considered for exploring levels of urbanization, due to their capability to capture the spectral variability that a modern urban landscape offers. Data and algorithms are two sides of a coin, while the data capture the variations, the algorithms provide suitable methods to extract relevant information. The literature reports a variety of algorithms for extraction of urban information from any given data, with varying accuracies. This article aims to explore the binary-classifier approach to target detection to extract certain features.

363 A Real-Time Photogrammetric System for Acquisition and Monitoring of Three-Dimensional Human Body Kinematics
Long Chen, Bo Wu, Yao Zhao, and Yuan Li

Real-time acquisition and analysis of three-dimensional (3D) human body kinematics are essential in many applications. In this article, we present a real-time photogrammetric system consisting of a stereo pair of red-green-blue (RGB) cameras. The system incorporates a multi-threaded and graphics processing unit (GPU)-accelerated solution for real-time extraction of 3D human kinematics.

375 Modeling Hyperhemispherical Points and Calibrating a Dual-Fish-Eye System for Close-Range Applications
Letícia Ferrari Castanheiro, Antonio Maria Garcia Tommaselli, Adilson Berveglieri, Mariana Batista Campos, and José Marcato Junior

Omnidirectional systems composed of two hyperhemispherical lenses (dual-fish-eye systems) are gaining popularity, but only a few works have studied suitable models for hyperhemispherical lenses and dual-fish-eye calibration. In addition, the effects of using points in the hyperhemispherical field of view in photogrammetric procedures have not been addressed. This article presents a comparative analysis of the fish-eye models (equidistant, equisolid-angle, stereographic, and orthogonal) for hyperhemispherical-lens and dual-fish-eye calibration techniques.
With more than 80 percent of its 1,190 coral islands standing less than 1 meter above sea level, the Maldives has the lowest terrain of any country in the world. This makes the archipelago in the Indian Ocean particularly vulnerable to sea level rise. With global sea level rising 3 to 4 millimeters per year, and that rate expected to rise in coming decades, some analysts anticipate a grim future for the Maldives and other low-lying islands. One study concluded that low-lying islands could become uninhabitable by 2050 as wave-driven flooding becomes more common and freshwater becomes limited. The Intergovernmental Panel on Climate Change anticipates sea level could rise by about half a meter by 2100 even if greenhouse gas emissions are sharply reduced or rise up to 1 meter if greenhouse gas emissions continue to increase strongly.

While the Maldives government has explored plans to purchase land on higher ground in other countries as an insurance policy against sea level rise, planners are also working to enhance the resilience of the country’s current islands. One example is Hulhumalé, a newly constructed artificial island northeast of the capital, Malé.

The pair of Landsat satellite images show just how much the area has changed between 1997 and 2020. Construction of the island, designed to relieve crowding in Malé, began in 1997 in a lagoon near the airport. Since then, the island has grown to cover 4 square kilometers, making it the fourth largest island in the Maldives. Hulhumalé’s population has swollen to more than 50,000 people, with 200,000 more expected to eventually move there.

The new island, built by pumping sand from the seafloor onto a submerged coral platform, rises about 2 meters above sea level, about twice as high as Malé. The extra height could make the island a refuge for Maldivians who are eventually driven off lower-lying islands due to rising seas. It could also prove to be an option for evacuations during future typhoons and storm surges.

Hulhumalé is not the only island in the Maldives that has seen major changes since the 1990s. Reclamation projects have enlarged several other atolls in similar ways in recent decades. Among them is Thilafushi, a lagoon to the west that has become a fast-growing landfill and a common location for trash fires (note the smoke plume blowing to the southwest in the 2020 image). Gulhifalhuea is the site of another land reclamation project that is opening up new manufacturing and industrial space.

There is one piece of positive news: natural processes on coral reef atolls (like those in the Maldives) might make the islands more resistant to sea level rise than their low elevations might initially suggest. Multiple studies, many of which use Landsat observations, show that most coral atoll islands in the Maldives and elsewhere have remained stable or even grown larger in recent decades.

Scientists are still studying why, but some research indicates that storms and floods that wash over islands can move offshore sediment onto the island surface, building the island up in the process. Other research shows that healthy coral reefs can grow upward even when seas are rising by producing abundant sediment.

“The key thing to understand is that these islands aren’t static. They don’t sit passively as if they were in a bathtub and slowly drowning,” said Murray Ford, a geologist at the University of Auckland. “They are constantly being reshaped by oceanographic and sedimentary processes.”

These natural processes may offer only limited protection to highly developed islands, partly because the construction of sea walls can disrupt the movement of sediment and human activity often degrades the health of coral reefs. “Once an island is on an engineered pathway, it can’t easily get off it. Islands that are being built on reclaimed land must factor in sea level rise and build higher off the ground,” said Murray. “For islands that are unpopulated, or sparsely populated, care should be taken to not interfere with the natural ability of islands to adjust to changes in sea level.”

For more information, visit https://landsat.visibleearth.nasa.gov/view.php?id=148158.

MAPPING KUWAIT OIL FIELDS WITH LEICA LIDAR AND IMAGING SENSOR SYSTEM

Majeed Al-Muwail, Faisal Al-Bous, Nasir Osman, Ramesh Mahishi Murthy
Operations Technical Services (South & East Kuwait), Kuwait Oil Company, Kuwait
Introduction

The Infrastructure Master Plan (IMP) section of the Operations Support Group in Kuwait Oil Company (KOC) is responsible for developing a Master Plan to manage all the oil field’s surface footprint. Recently, KOC made an important decision that would impact the quality and accuracy of the data used for planning, operations and Health Safety & Environment (HSE) activities. In order to achieve the purpose of supporting automation, coordination and management of the company’s geographic information and resources, IMP completed its first ever, aerial acquisition of high resolution digital images of 10 cm and LiDAR (Light Detection and Ranging) of 4 ppm. Orthophotos and LiDAR bare-earth Digital Elevation Model (DEM) will be produced at two different resolutions (10cm) and (20cm) using cutting-edge technologies.

Project Approach

Project planning involved numerous coincident activities. KOC selected the Area of Interest (AOI) based on the geographic distribution of KOC services/ facilities all over the State of Kuwait. KOC fully understands that the AOI includes some military areas and Kuwait international border lines that require special consideration. Coordination was exercised to obtain necessary permits from the Directorate General of Civil Aviation, Ministry of Defense and Air Force.

Distribution of ground control points (GCP) was planned and signalization was performed to support the accuracy requirements of the project.

Aerial Data Acquisition

The option of using Unmanned Aerial System (UAS) over a manned aircraft was first evaluated. Factors considered for the evaluation included:

1. Area of Interest: Spread over 8000² Km. It was not feasible to use a UAS.
2. Security: The AOI is close to international borders and includes vital defense installations and air bases, making it impossible to fly a UAS over them.
3. Permissions: Obtaining permissions from Directorate General of Civil Aviation (DGCA) to fly a UAS across half the country is next to impossible.
4. Safety: If there were a failure in the UAS like disconnection of GPS lock or a mechanical failure, the UAS could go rogue and cause safety concerns to KOC assets or violate the geo-fence.
5. Endurance: The endurance of a UAS is dependent on the weather, headwinds, etc. The maximum flight time for the UAS could be about an hour making it difficult to complete acquisition in one flying season.
6. Payload: Although the scope included acquiring both aerial images and LiDAR data, getting a UAS to support simultaneous collection was not possible.

Keeping above factors in mind, the choice of sensor was the second step towards aerial data acquisition. Since the aim was to acquire digital images and LiDAR data over the same area, flying two sorties was impractical. Therefore, Leica RCD 30 Digital Imagery Sensor integrated with Leica-70-HP-MPIA LiDAR platform with simultaneous collect facility was selected. RCD30 is the first medium format, single head camera which collects perfectly co-registered 80 MP RGBN multispectral imagery. Its innovative features and configuration flexibility support photogrammetric and remote sensing applications, offering performance otherwise only available from large format airborne sensors. Leica-70-HP LiDAR sensor has a unique, sophisticated, optical beam-splitting system that bifurcates the laser beam, resulting in an effective scan rate of 500,000 points per second. This gives the ALS-70-HP the ability to achieve the highest point density at a given altitude of any system currently in the market, resulting in exceptional acquisition efficiency (www.leica-geosystems.com).

The contractor was entrusted to choose an aircraft that could accommodate the sensor system, be able to fly steady and make tear-drop turns within short distances. Rockwell International 690A was an ideal choice for this. Pictures in Figure 1 are of the aircraft and equipment used for acquisition.

Area of Interest

Aerial acquisition was planned at 10 cm resolution digital images and 4 points per square meter of LiDAR data for highly developed areas, and 20 cm resolution digital images and 0.5 points per square meter of LiDAR data for the other areas. Area of interest towards North Kuwait had to be restricted to three (3) nautical miles from the Iraq International border.

Photogrammetric Engineering & Remote Sensing
0099-1112/21/313–317
© 2021 American Society for Photogrammetry and Remote Sensing
doi: 10.14358/PERS.87.5.313
Challenges during Aerial Acquisition

The industry experts would agree that aerial acquisition, though sounds easy, is quite a tedious task and influenced by various external factors. The experience of IMP was no exception. Some of the challenges faced during acquisition and mitigation steps taken were:

1. **Weather**: Kuwait experienced abrupt weather changes during 2019. Although the acquisition was planned to commence during the best season for aerial survey in Kuwait, i.e. November to February, the planned 19 days of acquisition actually took slightly over four (4) months due to unexpected strong winds making it difficult to stick to the planned flight path. Other challenges faced during acquisition were unexpected rains, haze, clouds, dust, and fog. This was overcome by extending the schedule and permissions assigned for aerial acquisition.

2. **Smoke**: Although the Production and Operations teams were informed prior to aerial acquisition to keep the flaring activity to minimum on the days of aerial photography, it was not always feasible due to operation and production constraints. Re-flights over the same areas were performed to get images with minimum smoke around production facilities.

3. **Turbulence**: One of the activities during production of oil is to burn gases that usually accompany oil through a smoke stack. This process is called as flaring. As part of production process, another activity is to burn out oil during asset maintenance. The air in the atmosphere heats up around the flare stacks or burn pits causing temperature and pressure variance. This makes the aircraft lose altitude around these areas. The air-crew were prepared well to control the aircraft around these areas, yet maintain the course and altitude manually. In order to maintain good quality data, the affected areas were re-flown.

4. **Flying close to international borders**: The area of interest for aerial acquisition included areas close to international borders that consist of crucial KOC assets. These areas thus involved close coordination with the Aviation and Defense authorities on both sides of the border and due care was taken not to violate any aviation laws. Experienced air crew were specially identified for the process who carefully piloted the aircraft through tight boundaries, yet covering a maximum AOI.

Image Data Processing

Data processing is a vital step for interpreting the results and drawing meaningful conclusions. Digital Aerial Triangulation (AT) was performed immediately upon the completion of aerial acquisition. The purpose of AT is to densify horizontal and vertical control from relatively few ground control points (GCPs). Since obtaining GCPs is a relatively significant expense in any mapping project, AT procedures were used to reduce the amount of field survey required by extending control to all stereo models.

The surveyed control, along with the reduced image coordinates, served as input into a combined block adjustment.
Three dimensional, simultaneous least squares adjustment by bundles, commonly referred to as “bundle adjustment”, was undertaken using Inpho Match-AT GPS adjustment software. A series of AT solutions were completed. The final adjustment and the optimal solution to be used for mapping, included all control points as constraints.

The AT results are being used for topographic data compilation and creation of Digital Terrain Model (DTM) by compiling breaklines photogrammetrically.

**LiDAR Data Processing**

Post processing of all LiDAR data flight, strips were completed to perform calibration and verify quality and coverage. In order to ensure a homogenous surface, adjustments were made to the orientation and/or linear deviation of individual and overlapping swaths to obtain the best fit relative accuracy. Project-wide calibration was evaluated using advanced vector matching analysis; and trajectory based solutions were applied. This procedure was repeated interactively until the residual errors of overlapping swath was reduced. A visual and statistical analysis was completed using elevation difference intensity raster and vector based accuracy reporting.

LiDAR data was then filtered and classified to separate terrain data from other data on land cover and manmade features. Only a ground/non-ground classification was performed for this project. A variety of commercial and proprietary software were used to build macros for automated classification. The macros were specific for the landscape of the project taking into consideration terrain relief, ground cover and natural and manmade features. The routines were used to classify points based upon the laser attributes including intensity, elevation and the numeric value of the returns.

**Orthophoto Creation**

Aerial images were ortho-rectified using the AT results and LiDAR DEM. Color balancing and mosaicking were performed to ensure the final imagery is both radiometrically and geometrically seamless. The orthophotos were partitioned and written out in compliance with the desired naming convention in GeoTIFF and MrSID format. The resulting orthophoto tiles were seamless with no overlaps or gaps between them.

On completion of data processing, KOC produced the following outputs to support planning, operations and HSE activities:

1. Digital Orthophotos: Images that are rectified to accurately represent the ground and could be used as photo maps.
2. Bare Earth DEM: Accurate representation of ground features or hypsography which could be used for generating contours depicting slopes.
3. 3-dimensional Vector Maps: Compiled photogrammetrically to represent the above ground features such as buildings, pipelines, utilities, etc., in 3 dimensions (X, Y, Z) accurately.
Quality Assurance and Quality Control

QA/QC is a process used to ensure the highest probability of creating quality products to predefined standards. KOC prepared a detailed measurable Acceptance Criteria to guide the production activities. It provides the basis for developing the supporting necessary tools to check the final deliverables and to confirm the all the delivered products have passed the established listed Acceptance Criteria. This includes in-house developed scripts to automate the checking (file naming convention, format, projection parameters etc.) as well as visual checks to confirm the adherence of the data to the predefined acceptance criteria.

Statistical and testing accuracy assessment has been implemented to verify conformance with accuracy requirements of the project. The positional accuracy is calculated statistically from well-known Ground Control and resulted in 10 cm for 10 cm orthophotos and 20 cm for that of 20 cm.

Conclusion

The anticipated challenges creating setbacks to the project as mentioned earlier were used to create and update the risk register and lessons learnt for future projects.

1. Although, using the imaging and LiDAR sensors for simultaneous data collection reduced the collection time, mobilizing two sets of aircraft and equipment would make the collection process faster, resulting in faster data output. Also, the re-flights of some areas could be avoided. It is understood that the process would be expensive, but the benefits reaped in getting the data faster far outweighs the costs incurred.

2. KOC field assets changes continuously. The data thus collected could be processed much faster resulting in usable information available to the field teams at a faster pace.

3. Periodic updates need be performed so that the information available to the field teams are up-to-date.

The initiative of KOC in creating high resolution orthophoto maps, LiDAR DEM, DTM and 3-D topographic maps shall yield the following benefits to the company:

1. Creation of orthophotos to serve as a base map supports in daily planning, operations and HSE functions. The challenges faced by the field survey team in day-to-day operations, such as route planning in terms of safest route, and avoiding hazardous areas, etc. are eliminated.

2. Visual reference for infrastructure planning provides a suitable platform to study the impact of KOC operations. One of the main functions of IMP is effective land management. With the use of high resolution orthophotos and LiDAR DEM, activities such as pipelines routes, land reservation, asset demarcation, removal of abandoned assets, etc. have become efficient.

3. Enhanced data collection in inaccessible and geo-hazard areas through 3D compilation.

4. Producing LiDAR data (Elevation Data, Terrain Datasets) to support:
   a. Orthophoto rectification.
   b. 3D simulations and spatial analysis.
   c. Hydrological modeling.
   d. Better representation of the terrain.
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.
This month’s column is a twist on the “standard” GIS Tips & Tricks and focuses on a highly technical area of photogrammetry, namely Aerial Triangulation and gives us a brief history of the technology. Dr. David Maune contributed this column and he opens up the “black box” for a little trickery that enables low-cost, high-precision imagery. Enjoy.

Today, Aerial Triangulation (AT) is performed with “black box” technology that most users don’t understand. My “trick” in teaching AT is to review all generations of photogrammetry that led to today’s digital photogrammetry and Structure from Motion (SfM).

**Fundamentals of Photogrammetry** — To reconstruct the 3-D geometry that existed when each aerial photo is taken, AT determines the 3-D ground-coordinate positions (X/Y/Z) of the camera focal point and angular orientation (ω/φ/κ) of the camera when each photo is taken, where: X is the x-coordinate (Easting), Y is the y-coordinate (Northing) and Z is the elevation of each photo’s focal point, e.g., o1 and o2 in Figure 1; omega (ω) is the roll around the x-axis, the direction of flight in the photo coordinate system; phi (φ) is the pitch around the y-axis, horizontally perpendicular to the x-axis; and kappa (κ) is the yaw around the z-axis, vertical and perpendicular to the x-axis and y-axis. Upper case letters (X/Y/Z) represent ground coordinates, typically in State Plane or UTM meters; lower case letters (x/y) represent photo coordinates, typically mm in a photo coordinate system.

At Figure 1, the two photos are shown in the position of film positives. Assume f is the camera focal length (e.g., 6” for older film cameras), o1 and o2 are the lens’ focal points for the camera when photos 1 and 2 were taken; x1 and y1 are the photo coordinates of point p1 on photo 1, and x2 and y2 are the photo coordinates of point p2 on photo 2. In theory, the two lines (light rays) drawn from o1 through p1 and from o2 through p2 should intersect at point P, enabling photogrammetrists to map the 3-D coordinates of point P on the ground. However, without correct position (X/Y/Z) and orientation (ω/φ/κ) of the two photographs, those two light rays will never intersect at point P; ground features cannot be focused in stereo, and terrain features cannot be mapped in 3-D. Without relative orientation (RO) between a pair of stereo images, analysts will see “parallax” in the stereo model. I think of parallax as the displacement amount in x and y at any point (e.g., point P) when the two lines do not intersect and are out of focus. Both x-parallax and y-parallax need to be removed for 3-D mapping from photogrammetry. Without absolute orientation (AO), the model will not be scaled and leveled to fit ground control.

**Analog Photogrammetry** — First generation analog photogrammetry was used from the 1930s into the 1970s. Multiplex and Kelsh plotters were the primary optical stereo plotters used, subsequently replaced by optical-mechanical analog stereo plotters. Optical and optical-mechanical analog plotters physically reconstructed 3-D stereo geometry, at reduced scale, from metric film cameras. Both RO and AO of analog stereo pairs were lengthy and demanding processes. Figure 2 shows the first two projectors in a string of many Multiplex projectors that replicated a flight line with each stereo pair of photos having approximately 60% forward overlap so that many features were imaged on three successive photos as shown in Figure 3. Reduction printers generated 2” × 2” glass diapositives from the original 9” × 9” film negatives. Then anaglyph (red/blue) glasses were used so that the photogrammetrist saw the left image with his/her left eye, and the right image with his/her right eye. But the stereo model could not clearly be seen

---

**Photogrammetric Engineering & Remote Sensing**
0099-1112/21/319–321
© 2021 American Society for Photogrammetry and Remote Sensing
doi: 10.14358/PERS.87.5.319
and focused until the two projectors were brought into RO by adjusting the height of the tracing table to remove x-parallax at six Von Gruber points (points “a” through “f” in Figure 3) and iterative adjustment of the $\omega/\phi/\kappa$ dials shown in Figure 2 to remove y-parallax at those points. This process could take 8 hours per stereo model. Subsequently, AO is required to adjust the relatively oriented stereo models to the proper map scale (by adjusting the baseline distance between stereo images (white arrow in Figure 2) and other adjustments to fit photo-control points and to level the models. After RO and AO, the map manuscript is plotted by moving the tracing table (with pencil beneath) while viewing the stereo model in 3-D. Contour lines were drawn by keeping the platen on the tracing table at the same height while moving, keeping the platen’s “floating point” on the ground while tracing. This required a great deal of training and human skill for compiling topographic maps.

Reference the red circles in Figure 2, y-parallax is cleared at each of the six Von Gruber points after first clearing the x-parallax at each point by adjusting the height of the platen on the tracing table. For the first stereo model, photogrammetrists cleared y-parallax at “b” with $\kappa$ at projector II; then they cleared y-parallax at “e” with $\kappa$ at projector I. Next they cleared y-parallax at “c” with $\phi$ at projector II and they cleared y-parallax at “f” with $\phi$ at projector I. Then they introduced one-half the y-parallax existing at point “a” or “d” in the opposite direction using $\omega$ of either projector. This process was repeated until no parallax existed at any of the six points. For subsequent stereo models, a similar process was used but with all adjustments to $\kappa$, $\phi$ and $\omega$ made only to projector III, holding projector II fixed from the prior relative orientation - then adjustments to Projector IV holding projector III fixed, etc.

Analytical Photogrammetry — Second generation analytical photogrammetry was dominant from the 1970s into the 1990s. PUG stereoscopic point transfer instruments were used to precisely drill small holes in the film emulsion for points in the triple-overlap region for pass-points and tie-points (see Figure 3) so the x and y coordinates of those points could be precisely measured in over-lapping and side-lapping images. Using collinearity equations, analytical plotters mathematically reconstructed the 3-D stereo geometry from metric film cameras. Part of the math for analytical photogrammetry included film camera “interior orientation” which defines image space coordinates based on camera calibration parameters such as precise focal length and lens distortion models inside the camera itself. These analytical models also applied corrections for atmospheric refraction and other parameters that could be mathematically modeled.

Least Squares adjustments, critical for analytical, digital and SfM photogrammetry, are used to solve an overdetermined system of equations (many more known values than unknown values) based on the principle of least squares of observation residuals. The “knowns” are survey control points, camera cal-
ibration parameters, and 2-D (x/y photo coordinate) measurements made on the pass points, tie points, and photo-identifiable target points to be mapped. The “unknowns” are the X/Y/Z and ω/φ/κ for each photo as well as the X/Y/Z ground coordinates of each pass point, tie point, and target point. The “trick” is to observe as many points on the ground as you can in the triple-overlap areas because each unknown point adds three “unknowns” (ground X/Y/Z coordinates) to the least squares solution and six “knowns” in triple-overlap areas where x/y photo coordinates are measured on each of three photos.

Digital Photogrammetry — Third generation digital photogrammetry was introduced in 1991 by the U.S. Army Engineer Topographic Laboratories (ETL) that developed the Terrain Information Extraction System (TIES). TIES included the first film scanner and the first digital stereo photogrammetric workstation (DSPW), using polarized glasses to see in stereo on a computer screen, and software (now marketed as SOCET SET) developed for the DSPW. With digital photogrammetry, DSPWs mathematically regenerate 3-D stereo geometry from digital images. The main “trick” from digital photogrammetry is that it relies on automated image correlation (image matching) to identify conjugate points on overlapping digital images and automatically generates thousands of pass points and tie points for large AT block adjustments. It also exploits the benefits of GPS and IMU technologies that provide excellent a priori estimates of position and orientation of each image so automated image correlation is efficient and least squares adjustment provides excellent a posteriori values for all unknowns. SOCET SET, and similar software packages by other firms, enables the production of digital orthophotos and automatically generates DEMs and digital feature data. Digital photogrammetry became more efficient in the early 2000’s when digital cameras were developed that added numerous AT benefits.

Structure from Motion (SfM) — Emerging in recent years, fourth generation SfM is essentially digital photogrammetry, using highly overlapping digital imagery from consumer-grade, non-metric cameras (Figure 4). SfM imagery is often acquired from unmanned aerial vehicles (UAVs). SfM software automatically identifies conjugate features in multiple overlapping images using image matching. These features are tracked from image to image and are used to calibrate cameras and estimate camera positions (X/Y/Z) and orientation (ω/φ/κ) of each image so that target features can be mapped on the ground with high accuracy at a fraction the cost of traditional photogrammetric techniques. There is no requirement in SfM for any a priori knowledge of the camera’s interior or exterior orientation or the scene geometry; these quantities are recovered empirically through a redundant network of matched feature points during the bundle block adjustment AT process. The ability to derive these quantities sets SfM apart from traditional techniques and allows for the use of lower cost cameras for which the calibration is unknown. Now that’s a smart “trick.”

Please feel free to share yours with us. Send your questions, comments, and tips to GISTT@ASPRS.org.

Dr. Dave Maune is Chief Scientist for the GTS group at Dewberry Engineers in Fairfax, VA. Maune has been an industry leader in photogrammetry, IFSAR and lidar and is the editor and principal author of all three editions of Digital Elevation Model Technologies and Applications: The DEM Users Manual, published by ASPRS. He was the ETL Director when TIES was developed. Dr. Al Karlin is with the Geospatial and Technology Services (GTS) group at Dewberry Engineers in Tampa, FL; as a Senior Geospatial Scientist, Karlin works with all aspects of lidar, remote sensing, photogrammetry, and GIS-related projects.
Cartography is an illustrated and graphically rich compilation of map design principles. It acts as a reference, exemplar, and reminder of how we present geospatial data to a wide range of end-users. Unlike a traditional text or reference book, it is laid out with topics in alphabetical order - essentially a dictionary of cartographic principles. Each topic is condensed into a two-page spread, often one page of text and an accompanying illustration. In between are interspersed cartographic vignettes - examples and interesting cartographic approaches.

I found this presentation style encouraged random exploration of cartographic design ideas, appropriate to the book’s intent. It is less a ‘how to make maps’ and more a compilation of principles that might help in your map-building process. These range from color mixing to presentation of different data types, to making maps ‘fun.’ It also includes advice to mapmakers: traps to avoid, ways in which your mapping efforts may be read unintentionally, expressing your ideas visually to a possibly as-yet-unknown audience.

I have had this book for quite some time by the time of this review, and still find myself discovering more to read. By design, it is a book for sampling in small chunks over a long time rather than reading end to end. For me it was easy to read and engage with, although I needed to keep handy a way to look up a number of terms specific to the fields of technical/academic cartography. If readers are not familiar with design principles (color, typography, etc), they may need to spend some time in dictionaries. However, there are very few places where deeply technical terms are not explained thoughtfully. I also marvel at the succinct expression of this book. It is an incredible effort to condense these topics so well, truly reflecting the authors’ depth of experience.

I found myself disagreeing with the author’s perception of a ‘good map’ at times. However, this is to be expected and encouraged! The author spends significant space on fitting maps to our own contexts, trying (but not always quite succeeding) to avoid distinction around ‘good’ or ‘had’ maps. This, however, is difficult because when one starts thinking about the design principles and end uses of a map, there truly are terrible ways to present information visually. If there is anything to improve on for future editions of Cartography, it is only the unavoidable cultural lens through which it is written. Here I refer readers to Britta Ricker’s extensive 2020 review (https://doi.org/10.1080/23729333.2020.1711547), and add only that for me, it veered at times to a highly westernised perception of ‘good design’ (a criticism I must level at myself also).

To summarise, Cartography is a welcome addition to my bookshelf and I will continue to explore it as I come across design decisions in my data presentations. I would recommend this book to anyone who is interested in better visual communication of geospatial data on a map - from professional cartographers to students, designers and researchers. After all, geospatial research is also a communication job! I also hope readers find a cunning easter egg as they flip through its pages.
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’s reprinting prior articles from the column. This month’s article on Jamaica was originally printed in 2002 but contains updates to their coordinate system since then.

Inhabited by Arawak Indians when Columbus discovered the island in 1494, a Spanish colony was established in 1509. The island was captured by British naval officers William Penn and Robert Venables in 1655, the Spanish being finally expelled in 1658. At that time the total population of the island was estimated at 3,000, chiefly the slaves of the eight hidalgos who were the “lords proprietors,” opposed to further European immigration. By then, the native Indian population had been wiped out by the conquistadors, just as the Arawak previously had done to the other peaceful race – the Lucayan of the Bahamas. Of the indigenous population nothing remains but the name of the island, “Xayamaca,” the “Land of Springs,” and numerous examples of native craft that were left in caves. Jamaica and the Cayman Islands (PE&RS, November 1998) were ceded to Britain in 1670 under the Treaty of Madrid and Jamaica remained a British Colony until it became independent in 1962 within the Commonwealth. The island of Jamaica is slightly smaller than Connecticut, and it has a 1,022-km coastline. Jamaica claims a 12-nautical-mile territorial sea, and a continental shelf of 200 nautical miles or to the edge of the continental margin. The terrain is mostly mountainous with a narrow discontinuous coastal plain. The lowest point is the Caribbean Sea and the highest point is Blue Mountain Peak (2,256 meters).

“The island does not lend itself to camping except in the Blue Mountains, so one of the bugbears of the profession is the long walks to and from the field each day. This fact is best illustrated by the experiences of surveyors working in the ‘Cockpits,’ who have found their days to be each divided into three parts – three hours for the walk to work over trails which have such picturesque names as the ‘Devil’s Staircase,’ leading through the ‘District of Look Behind,’ six hours behind the instrument and four hours for the return journey to his car or to his lodging” (Survey Review, 1955).

The earliest topographic maps of Jamaica were four sheets of Kingston and of Lower St. Andrew at the scale of 2½ inches to the mile made by the Military Authorities in 1909-1912 and known locally as the Pomeroy Map. All other maps were based on compass surveys until the late 1930s.

The Department of Lands and Surveys was formed in 1938, and “the layout of the trigonometrical survey was complete by the end of the year. All lines had been cleared; at the 38 primary stations, 4-foot concrete piers had been built and somewhat smaller piers at 20 secondary stations. At one-third of the primary stations, the observations were completed with the two geodetic Tavistock theodolites in use; they showed triangular closures averaging about 1” (Tavistock was the name of the town where the manufacturer was located. The instruments were actually made by Cooke, Troughton and Simms – CJM). All angular observations were made between sunset and midnight to projectors; sixteen pointings on eight zeros exhibited an average range of 4”. It was hoped to com-
complete the measurement of a 3½-mile base in Westmorland by January 1939. Invar wires were to be used. An azimuth was observed in Trelawny, with a probable error of about ± 1°. The fixing of latitude, longitude and check azimuths had not been undertaken at the date of reporting." “The azimuth observations were made at Maxwell Hall hill, Kemshot (sic) on the 28th and 30th of March, 1938. Using Ursa Minor at or near elongation the azimuth of the side Maxwell Hall-Etingdon was determined. In all, thirty-two pointings were made, and the probable error of the result is 1°. The azimuth of the side was calculated (and later corrected) to be α₀ = 81° 16′ 36.25″.”

“Datum Latitude and Longitude: — In accordance with the original instructions from the Geographical Section (of) the General Staff, the Flagstaff in Fort Charles, Port Royal, the latitude and longitude of which had previously been determined to be Φ₀ = 17° 55′ 55.8″ N, Λ₀ = 76° 50′ 37.26″ West of Greenwich, was taken as the triangulation origin. It has been connected to the triangulation by intersections from: Wareika, Montpelier, Red Hills, Rodney’s Lookout, Plumb Point.”

“Datum Level: — A line of level (sic) was run from the Plumb Point station to the Admiralty Datum in the swimming bath at Port Royal. The mark is reported to be … 3.50 ft. above mean sea level. The distance from Plumb Point to the Port royal bath is 4½ miles and the disagreement between the outward and return leveling was .086 of a foot, a very good result obtained by Sergeant D.K. Black, R.E. (I. J. Harris, Capt., R.E., 20th May, 1939).” Note that ordinary differential leveling would allow a total error of ± 0.106 feet even by 21st Century standards.

The Clarke 1880 ellipsoid was referenced (and modified for Jamaica) such that a = 6,378,249.136 m and 1/f = 293.46631. The Lambert Conical Orthomorphic was chosen as the projection for Jamaica, and quite interestingly was designed to be a tangent conic where m₀ = 1.0. Using feet as the unit of measure until 1980, the defining parameters of the “Jamaica Foot Grid” are central meridian, λ₀ = 77° W, latitude of origin, φ₀ = 18° N, False Easting = 550,000 feet, and False Northing at the latitude of origin = 400,000 feet, all on the Port Charles Flagstaff Datum of 1938. The first complete coverage of the island with aerial photography was flown in 1941-1942. The British Directorate of Overseas Surveys (DOS) compiled 12 topographic maps of the island at 1:50,000 scale in 1947, and this was the first DOS project undertaken by the new agency. Brigadier Martine Hotine (the father of the Rectified Skew Orthomorphic Oblique Mercator projection) was the only Director of the DOS for its entire existence. Of course, the units were feet for this edition, even though the scale was metric.

In April of 1969, Keith A. Lee wrote: “In the original scheme only one azimuth had been observed by the Royal Engineers; that line Maxwell Hall to Etingdon being situated near the western end of the island. It was considered prudent to determine the azimuth of another line and preferably one in the eastern part of the island. For reasons of accessibility and convenience the line Coopers Hill to Nutfield was chosen. On the night 24/25 October, 1968, the azimuth of this line was deter-
mined by observations on Polaris. The instrument used was Wild T3 No. 53112 and the observations were made on the 16 circle positions recommended in “The Retriangulation of Great Britain.” Not having the equipment for a proper determination of astronomical coordinates, the listed geodetic position of the stations was used in computing the azimuth. It was noted that this computed azimuth was 10.01” larger than the listed geodetic azimuth of the Royal Engineers. The observation was thought to be of a fairly high standard – the p.e. being ± 0.29” – and although a systematic swing was present in the individual determinations it was certainly not of the order of 10”. Further investigations by Lee proved that the original reference azimuth of the old Fort Charles Flagstaff Datum of 1938 was in error as a combination of mathematical blunder and probably deflection of the vertical. Note that the reference azimuth quoted with the definition of the old datum was the corrected value that was carried in the readjustment discussed in the following paragraphs. Lee did recommend that the original origin point for the datum be retained, and it was.

From Watson almost 40 years after Captain Harris: “In 1969 it was decided to readjust the primary triangulation of Jamaica and to incorporate into the new adjustment all additional and relevant observations made since the previous adjustment. In particular, many electronic distance measurements and some additional astronomical azimuth information were now to be included. At the same time it was decided to change spheroids and carry out all computations on the Clarke 1866 spheroid instead of the current Clarke 1880 spheroid! These decisions were agreed between the Jamaica Survey Department and the Directorate of Overseas Surveys. Although quite unusual when readjusting a datum to go to an older ellipsoid of reference, in this case the reason was that the Inter-American Geodetic Survey had extended the North American Datum of 1927 into the Caribbean. In the case of Hong Kong (PE&RS, January 1998), the ellipsoid was changed to an older one to allow for a ‘better fit’ for a readjustment.

The bulk of the triangulation was observed and adjusted by a team of Royal Engineers during the period 1937-45. Finally, during 1951-62 a team of surveyors from the Directorate of Overseas Surveys (DOS), assisted by members from the local department, measured by Tellurometer a good selection for the primary lines” (Tellurometer was a brand name of electronic distance meter once manufactured in South Africa. – CJM).

“The local department would undertake the daunting task of producing abstracts and/or Photostat copies of all observed angles and lengths held by them. These abstracts together with the observed angles and lengths held by DOS would constitute the data for the new adjustment. Historically, this task involved using angles observed between 1937 and 1970 by the local department on different types of survey using different types of instruments. In addition, from 1960 onwards, length measurements by Tellurometer and other electronic distance measuring equipment were carried out by both DOS and the local department. DOS observed Tellurometer traverses all around the island between primary stations and produced a network of Tellurometer traverses in the Cockpit part of the country. The local department extended and strengthened existing minor work all over the island by triangulation and Tellurometer traverse.

During the next few years, the local department prepared and sent batches of data and annotated diagrams on map sheets. In 1971-72, after consultation with the local department and DOS, a detachment of Royal Engineers, under the code name “Calypso Hop,” observed additional angles and lengths in some of the weaker points of the framework. Finally in 1974, a start was made on the secondary and minor adjustment by producing fresh diagrams showing all the information supplied by the local department and held by DOS. Missing data, as revealed by the diagram, and obvious errors which arose during the plotting of the data, were referred back to Jamaica, which department continued to send further information as their diligent searches of their records brought them to light. Close and friendly cooperation with the local department helped considerably with this huge task of collating and collecting the data.

Naturally the moment had to come when a halt had to be called to any further additions to the data to be adjusted. Any weaknesses in the framework would be revealed in the results of the adjustment, and these weaknesses could be strengthened in the future, though, here, it must be added that serious consideration should be given to trying to hold adjusted coordinates for say 20 to 30 years before attempting a complete re-adjustment. Too much localized chopping and changing leads to confusion and a lack of continuity. This may appear to be a rather pragmatic approach but surveyors and engineers are pragmatic people and surveys have still to be started and computed from the best results available at the time.

It must be remembered that no coordinates are ever final, and, if certain small areas of the adjustment are weaker than the mainly strong whole, additional field work can be carried out in the future to strengthen those weaker areas and if absolutely necessary, small controlled readjustments carried out. Following on from the adjustments, some 291 offset points were fixed by azimuth and Tellurometer length and, finally, 36 additional points were computed mainly by the method of intersection. Geographical coordinates and Lambert Conical Orthomorphic projection coordinates were computed for 1,392 points” (W. Watson, April 1977).

The Jamaica Datum of 1969 retained the original origin point at the Fort Charles Flagstaff where $\Phi_0 = 17^\circ 55.558^\prime N$ and $\Lambda_0 = 76^\circ 50.26^\prime W$, the reference azimuth $\alpha_0 = 81^\circ 16.36^\prime 25^\prime W$, was used as corrected by Lee, and the ellipsoid of reference was now the Clarke 1866 where:

\[
\begin{align*}
R & = 6,378,206.4 \text{ m} \\
R & = 6,356,583.6 \text{ m}
\end{align*}
\]

The new unit of measure was the meter, the redefined “Jamaica Metre Grid” retained the original parameters except for the change in ellipsoid, the False Easting was now 250 km, and the False Northing at the latitude of origin was now 150 km. Deflection of the vertical at the flagstaff was equated to zero, as was the geoid-ellipsoid separation. However, the new grid did not appear on official topographic maps until 1981.

Control densification work continued by the Jamaica Survey Department, and DOS with its library was incorporated...
into the Ordnance Survey of Great Britain. A combined adjustment of the original 1969 DOS data along with the control densification and trigonometric levels was performed in 1984 with excellent results. Minor problems with the original 1969 adjustment were corrected, and the final adjustment was performed in blocks with a PDP 11/34 mini computer (that’s the same model of computer many of us old Photogrammetrists used to have that powered our analytical stereoplotters).

On 16 November 1994, the Government of Jamaica signed a formal treaty with the Government of the Republic of Cuba on the delimitation of the maritime boundary between the two states. Based on the mutually agreed principle of the equidistance method, a list of 106 points was filed with the Secretary General of the United Nations. Although the points were plotted on navigation charts, it was explicitly noted in the treaty that the charts were intended for illustrative purposes only, and the lines connecting the tabulated points were defined as geodesics on the North American Datum of 1927 and the Clarke 1988 ellipsoid (a geodesic in this case is defined as the shortest distance on the surface of the ellipsoid of revolution).

Jamaica filed a formal declaration with the Secretary General of the United Nations under the United Nations Convention on the Law of the Sea regarding its claim to territorial waters based on Jamaica’s archipelagic basepoints as of 14 October 1996. The list of coordinates is comprised of 28 points described in latitude and longitude to the closest integer arc second “referenced to the North American Datum of 1927 (NAD27) and based on Clarke’s (1866) spheroid with a semimajor axis of 6,378,206.4 metres and a flattening of 1/294.978.” Since the text of the declaration includes reference to those basepoints plotted on navigation charts, the projection is a normal Mercator and therefore the “straight lines” connecting the basepoints are ellipsoidal loxodromes or rhumb lines. Readers may recall that it was Snellius, the father of geodesy that coined the word “loxodrome” (PE&RS, February 2003).

NIMA lists the transformation parameters for Jamaica from the NAD27 to WGS84 as a mean of 15 stations observed throughout the northern portion of the Caribbean as: \( \Delta a = -69.4, \Delta \lambda = 0.37264639, \Delta X = -3 \text{ m } \pm 3 \text{ m}, \Delta Y = +142 \text{ m } \pm 12 \text{ m}, \Delta Z = +183 \text{ m } \pm 12 \text{ m}. \) In addition, in 1996 the U.S. National Geodetic Survey (NGS) observed several existing triangulation stations at the Norman Manley International Airport in Kingston, Jamaica. Dave Doyle of the NGS graciously provided the local coordinates of a number of points at the airport. In particular for “Airport 5,” \( \varphi = 17^\circ 56’ 06.776” \text{ N}, \lambda = 76^\circ 47’ 44.836” \text{ W}, \) Northing = 142,844,681 m, Easting = 271,626,526 m, and the observed NAD83 geodetic coordinates are \( \varphi = 17^\circ 56’ 16.27529” \text{ N and } \lambda = 76^\circ 47’ 41.29986” \text{ W}. \) Computing the three parameter geocentric transformation from JAD69 to NAD83 for this single point yields \( \Delta X = +58 \text{ m}, \Delta Y = +209 \text{ m}, \text{ and } \Delta Z = +392 \text{ m}. \) The large differences with NIMA mean values, however, have a simple explanation. NIMA values for Jamaica are for the transformation from NAD27 to WGS84; they are not from JAD69 to WGS84. Although both classical datums are referenced to the Clarke 1866 ellipsoid, note that the origin coordinates for JAD69 are exactly the same as for the Jamaica Datum of 1938: it’s at the Fort Charles Flagstaff! The reader will recall that the origin for NAD27 is at Meades Ranch in Kansas, and that is quite a long distance from Jamaica.

Prof. Glendon G. Newsome of the University of Technology in Kingston and Prof. Bruce R. Harvey of the University of New South Wales wrote a paper on transformations in Jamaica and they used four fiducial points on the island. The three parameters from JAD69 to WGS84 they solved for are \( \Delta X = +65.33 \text{ m } \pm 0.96 \text{ m}, \Delta Y = +212.46 \text{ m } \pm 1.49 \text{ m}, \text{ and } \Delta Z = +387.63 \text{ m } \pm 0.69 \text{ m}. \) Newsome and Harvey acknowledge the critical need for a good geoid for Jamaica, and they hope for the day when a new national datum may be established. Thanks also go to Prof. Hugart Brown, retired from the Metro State College of Denver, for his kind help over the years with Jamaican survey history. Thanks to Russell Fox of the Ordnance Survey of the United Kingdom and his soon to be closed International Library for all of his help.

### Jamaica Update

In 2008, the Ministry of Agriculture established a GPS RTK Virtual Reference System for the island consisting of a five-station network.


In 2012, The Jamaica VRS and Cadastral Surveying was discussed by Prof. G. G. Newsome, Prof. G. Peake, and Mr. R. Douglas in the September issue of Coordinates magazine.


In 2015, The Spatial Active Global Geomatrix was presented by Mr. Siburn Clark, RICS and Mr. Douglas Nelson at the 4th Annual Caribbean Valuation & Construction Conference in which they discussed the 13 CORS sites in Jamaica of the VRS system.

Can you briefly tell me about your background and your connection to ISPRS?
I am currently a PhD student at United Nations University – Institute for the Advanced Study of Sustainability in Tokyo, Japan. My background is in the fields of geodesy and remote sensing, as well as sustainability science. I am the President of the ISPRS Student Consortium, the representation of the youth to ISPRS. I lead the organization and coordination of events designed for students and young professionals.

What are a few connections between ISPRS and ASPRS?
I could name a few similarities, beginning with the promotion of remote sensing and photogrammetry. I think many of the activities and initiatives of both organizations are aligned in terms of contributing to the progress of our profession and the scientific community at large. Many key people in ISPRS from North America are also affiliated with ASPRS, such as Dr. Charles Toth, the 2nd Vice President and also the liaison between ISPRS and ISPRS SC, and Dr. Marguerite Madden, who has been supporting the Consortium immensely through the ISPRS Foundation.

Can you briefly explain or discuss the ISPRS Commissions?
ISPRS is one of the biggest international organizations in remote sensing, photogrammetry and spatial information science and the Technical Commissions (TCs) were created in order to execute the scientific and technical work of the Society. There are five different Commissions: Sensors and Systems (TC I), Photogrammetry (TC II), Remote Sensing (TC III), Spatial Information Science (TC IV) and Education and Outreach (TC V).

Who can join ISPRS? If so, what are the average costs for students and individuals?
Anyone who is interested in photogrammetry and remote sensing can join ISPRS, either through their respective national or regional organizations or as an individual member. The ISPRS Student Consortium is the organization within ISPRS that was created for the youth, and we also have different types of membership, including individual membership. The Individual Membership is free for both ISPRS and ISPRS SC.

What are a few benefits offered to members of ISPRS?
In general, if you are a member of the ISPRS and ISPRS SC, you have access to the professional network of the organization, opportunities to attend scientific gatherings and trainings, reduced fees for students and opportunities to apply for travel grants. Ordinary Members in ISPRS can host a Technical Commission and symposium, and vote in the General Assembly.
Are your member benefits different for students?
Yes, ISPRS and ISPRS SC benefits are different. Individual members in ISPRS SC have the right to vote in the General Assembly and can be nominated as a member of the Board of Directors, if eligible.

Can you tell me more about the ISPRS Keynote Speaker Programme?
The Keynote Speaker Programme is another opportunity for Ordinary, Sustaining and Regional Members of ISPRS to tap into the experts’ pool of ISPRS and to receive additional funding support for events. They can request for an international expert for their event to be sponsored by ISPRS.

What are a few of your educational and training programs that may be of interest to students?
One of the major activities of the ISPRS SC is the summer schools, which we host about 3 – 4 times a year in different countries. These summer schools provide about 3 to 5 days of lectures, practical sessions, socials and field trips for participants at a minimum cost. We also have the Webinar Series, where we invite experts in different topics of interest in the fields of remote sensing, photogrammetry and spatial information science to give a formal lecture and a Q&A session with the participants. The Virtual Rooms is another event that we host online in the form of a meeting and encourages more interaction between the speaker and the audience. We also publish the SpecTrum, our official Newsletter, where we feature interesting articles and interviews with experts in the profession.

Do you have any previous events recorded or any upcoming events?
The ISPRS SC has a YouTube channel that keeps a record of our virtual events, particularly for our Webinar Series and the Virtual Rooms. We also kept the recordings of the first virtual summer school last year. You can visit this link to check out these videos: https://www.youtube.com/channel/UCcdvYFJEwf1cGncArKckDqg.

One of the upcoming events of ISPRS is the second virtual edition of the ISPRS Congress to be held from July 4 – 10. The ISPRS SC also organizes the monthly GeoMixer events together with the Ladies of Landsat, Sisters of SAR, ASPRS SAC, IEEE-GRSS IDEA and Mato Gross do Sul Brazilian Student Chapter and the Asian Association on Remote Sensing (AARS). We have summer schools and webinars coming this year but details are yet to be finalized so stay tuned on our website and social media.

Where do you see ISPRS going in the future?
I think ISPRS is going towards increased collaboration and involvement of the youth in the very near future. In order to address many of the pressing global issues, working with other disciplines will be crucial in the coming years and the role of remote sensing, photogrammetry and spatial information science can further be strengthened by international societies such as ISPRS. The changes that we experienced during this pandemic also calls for recognizing the contributions of the youth and supporting them to become leaders of the future.

Is there anything you would like to tell me about ISPRS that I did not ask you?
You can visit the official websites of ISPRS (https://www.isprs.org/default.aspx) and ISPRS SC (http://sc.isprs.org/) for more information about our organization and details about upcoming events.

Please follow the ISPRS SC on Facebook (https://www.facebook.com/groups/isprssc/) and Twitter (@ISPRS_SC) for more updates!
NEW ASPRS MEMBERS

ASPRS would like to welcome the following new members!

At Large
Amr Ashraf Abdelraouf
Eshrat Fatima
Mahmoud Abdelatwab Mohamed
Lu Xiaoyan

Cascadia
Benjamin Rose
Joetta Zablotney

Eastern Great Lakes
Prof. Tao Liu
Rami Tamimi
Michael Evan Vega

Florida
Iarelis Hall
Murali Nair
Jeffrey Young

Gulf South
Ibraheem Ali
Madison Bradley
Michael Corley
James Bartholomew Gillis, PLS
Jose Abel Pilartes Congo, Jr.
Neeraj Sirdeshmukh

Heartland
David Holm, CP
Erica Kriner
Mike Madden

Mid South
Olusola-Ige Adetoro, PhD
Sabien Lindo
Katherine Markham
Steffen Matthews
Owen Smith
Stephen Todd
Tommy Lee White

North Atlantic
Liubov Dumarevskaya
Durga Prasad Joshi
Jason Parent, PhD
Harshana Maduranga Wedagedara

Northeastern
Liam Bhajan
Hamdy Elsayed
Ben Fraser
Shahriar Shah Heydari

Pacific Southwest
Marie Bouffard
Jeffrey Campbell
Alan De Nova
Paul Christian Dunn
Madison Hays Hernandez
Kelly James
Cyrus Khambatta
Kimberly Mauch
Ryan Porras
Everett Tien-Fong Spackman
Mike Stys
Lorraine Tighe, PhD
Megan Singleton, PhD

Potomac
Lauren Kailey Ballard
Ivy Lynn Knapp
Bruce A. Lillegard

Rocky Mountain
Amanda Aragon
James Collins
Joseph Kerski, PhD
Thomas J. Rokita
Loren Russell
Toby Terpstra
Glen W. Thurow

Western Great Lakes
Colby Borchetta
Thomas Lake

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

Your Path To Success In The Geospatial Community
Remote Sensing Monitoring for Urban Environment


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

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

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

Important Dates (Tentative)
- March 1, 2021—Submission system opening
- September 31, 2021—Submission system closing
- Planned publication date is December 2021

Guest Editors

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

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

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

Haiyan Yao, Ziying Li, Yang Han, Haofang Niu, Tianyi Hao, and Yuyu Zhou

Abstract
In vegetation remote sensing, the apparent radiation of the vegetation canopy is often combined with three components derived from different parts of vegetation that have different production mechanisms and optical properties: volume scattering $L_{vol}$, polarized light $I_{pol}$, and chlorophyll fluorescence ChlF. The chlorophyll fluorescence plays a very important role in vegetation remote sensing, and the polarization information in vegetation remote sensing has become an effective way to characterize the physical characteristics of vegetation. This study analyzes the difference between these three types of radiation flux and utilizes polarization radiation to separate them from the apparent radiation of the vegetation canopy. Specifically, solar-induced chlorophyll fluorescence is extracted from vegetation canopy radiation data using standard Fraunhofer-line discrimination. The results show that polarization measurements can quantitatively separate $L_{vol}$, $I_{pol}$, and ChlF and extract the solar-induced chlorophyll fluorescence. This study improves our understanding of the light-scattering properties of vegetation canopies and provides insights for developing building models and research algorithms.

Introduction
In many studies on vegetation remote sensing, sensors are directly employed to develop algorithms and models of the apparent reflected radiation (Grant et al. 1987; Tyo et al. 2006; Zhang et al. 2020). With the increased demand for quantitative and accurate remote-sensing applications, many studies have emphasized the need to distinguish between apparent radiation reflected from vegetation and radiation scattered within the vegetation (Brakke 1994; Umeyama and Godin 2004; X. Liu et al. 2019). Previous remote-sensing measurements of global vegetation have predominantly focused on solar reflected radiance in the range of approximately 350–2500 nm (Meroni et al. 2009). However, radiation sensors will record not only the radiation reflected by the vegetation surface but also that emitted by chlorophyll, which is termed chlorophyll fluorescence (Huang et al. 2013). Thus, neglecting the effects of atmospheric and remote-sensing sensors, radiation detected from the surface of vegetation within a spectral range of 350–2500 nm actually consists of three components: volume scattering $L_{vol}$, chlorophyll fluorescence (ChlF), and polarized light $I_{pol}$ (Huang et al. 2013). Although these components are combined into one measurement by the sensor, they represent different information about the vegetation. Specifically, these three radiant fluxes are derived from different parts of the vegetation, have different optical properties, and are generated by different processes (Bousquet et al. 2005).

Volume scattering, which describes light scattered by the leaf interior, involves the leaf constituents and internal structure and accounts for the vast majority of the reflectance characteristic of green vegetation (Rondeaux and Herman 1991). Due to multiple scattering within the leaf, volume scattering is diffuse and nonpolarized. The ChlF flux, which is emitted by chlorophyll, is typically obscured by $I_{pol}$, because the emission signal only accounts for less than 2% of the total absorbed light (Meroni et al. 2010; Hu et al. 2018). However, the ChlF intensity at specific spectral bands, such as the dark lines resulting from atmospheric O$_2$ absorption, makes a significant contribution to the corresponding reflectance (Kharuk and Yegorov 1990). Similar to volume scattering, ChlF also tends to be nonpolarized (Gillerson et al. 2006). Although ChlF accounts for a small proportion of the total reflected radiation, it plays a crucial role in vegetation detection. For example, it is an important indicator of photosynthesis (Fournier et al. 2012; L. Liu and Liu 2018). The intensity of ChlF radiation under natural light conditions is dominated by photosynthetically active radiation absorbed by vegetation; therefore, it can be used to assess the gross primary productivity of vegetation photosynthesis (Damm et al. 2010; Frankenber et al. 2011; X. Liu et al. 2019).

The final radiation component is polarized light, which is reflected from the canopy surface and indicates leaf surface properties such as optical roughness (Rondeaux and Herman 1991). As well as describing the reflection intensity, $I_{pol}$ is a light-wave signal commonly used in vegetation optical remote sensing to convey polarization information (Talmon and Curran 1986). Such information can effectively characterize the physical characteristics of the surface, distinguish vegetation types, estimate the biomass of vegetation, and evaluate the leaf tilt angle (Curran 1981; Vanderbilt et al. 1985). In addition, the geometric morphological characteristics of the vegetation canopy will affect its polarization and reflection characteristics (Rondeaux and Herman 1991). Specular reflection generated by the leaves is the main factor affecting the polarization characteristics of the vegetation canopy, and the direction of the leaves will change the degree and intensity of polarization reflection (Woessner and Hapke 1987; Rondeaux and Herman 1991; Litvinov et al. 2011). Previous studies have explained why the vegetation canopy produces polarization information, as well as the importance of polarization information in vegetation remote-sensing research (Huang 2019).
et al. 2013). Polarization information therefore provides an important reference for the development of vegetation remote sensing. It can be used in differentiating vegetation types, estimating the biomass of vegetation, and evaluating the leaf tilt angle of the vegetation canopy (Curran 1981; Vanderbilt et al. 1985).

A commonly used method of extracting chlorophyll fluorescence is the Fraunhofer-line discrimination (FLD) principle (Tyo et al. 2006; X. Liu and Liu 2015). Plascyk first introduced the method to extract solar-induced fluorescence from observed vegetation-reflected radiance (Tyo et al. 2006). Various studies have demonstrated the possibility of measuring solar-induced fluorescence using Fraunhofer lines or atmospheric absorption bands (McFarlane et al. 1980; Carter et al. 1996; L.-Y. Liu et al. 2006). The FLD method of extracting chlorophyll fluorescence commonly employs the standard FLD (sFLD) algorithm. Three-channel FLD (Maier and Günther 2001) assumes that the variation of reflectance and chlorophyll fluorescence is linear, and improved FLD (Alonso et al. 2007) uses two correction coefficients to demonstrate variations in reflectance and chlorophyll fluorescence.

Currently, polarization techniques are rarely used in vegetation-fluorescence inversion. Our team wants to use this inversion model to explore the possibility of polarization inversion of fluorescence and the accuracy of fluorescence inversion. The method used in this article is roughly similar to the FLD method, but we use polarization measurement for the ChlF inversion. Polarized reflection is a common phenomenon in vegetation remote sensing. The use of polarization measurement can open up a new idea of chlorophyll-fluorescence inversion. In previous research, $L_{pol}$ and ChlF have been assumed to be negligible or unmeasurable and ignorable (Grant et al. 1987; Woessner and Hapke 1987; Tyo et al. 2006). Therefore, the purpose of this study was to use polarization technology to extract vegetation ChlF by first separating $L_{pol}$ from the unpolarized portion $L_{unp}$ and then extracting ChlF from $L_{unp}$. The accuracy of this fluorescence extraction technique is verified using the SCOPE (Soil Canopy Observation, Photochemistry and Energy Fluxes) model (Hu et al. 2018).

**Methodology**

**Principle**

Due to absorption by the atmospheres of the sun and Earth, the spectrum of solar irradiance reaching the Earth's surface contains many small dark lines, termed Fraunhofer dark lines. Chlorophyll fluorescence can fill Fraunhofer absorption wells; thus, by comparing the intensity of a Fraunhofer line and its adjacent spectral region in the solar incident irradiance spectrum and vegetation-reflected irradiance spectrum, the FLD algorithm can be used to calculate the solar-induced chlorophyll fluorescence in two absorption bands. Fraunhofer lines provide a potential compensation method for vegetation-fluorescence detection under natural lighting conditions (L.-Y. Liu et al. 2006). This study employed the sFLD method, which utilizes the dark lines of solar atmospheric absorption and the Earth's atmospheric absorption; the two most commonly used absorption lines are $O_2-A$ and $O_2-B$. The principle of sFLD is shown in Figure 1. The wavelengths of $O_2-A$ and $O_2-B$ are 761 and 688 nm, respectively. The unaffected reference band on the left side of the absorption dark line is 756 nm for $O_2-A$ and 684 nm for $O_2-B$.

This study employed a method of vegetation-fluorescence inversion based on polarization technology, which can be divided into two main steps. First, Epipremnum aureum is used as the target vegetation. The apparent radiation signal of the Epipremnum aureum canopy is separated into $L_{pol}$ and $L_{unpol}$ using polarization technology. Second, sFLD is used to distinguish $L_{pol}$ inside the blade and ChlF from $L_{unpol}$. A flowchart of the chlorophyll-fluorescence inversion method is shown in Figure 2, where $F$ is the chlorophyll-fluorescence value obtained from the algorithm.

To separate the polarization components (Wu and Zhao 2005), polarization technology should be used to pass the Stokes vectors ($I$, $Q$, $U$, $V$), where $I$ represents the intensity of the light, $Q$ represents the size of the semimajor axis of the polarization ellipse, and $U$ represents the direction of the semiminor axis of the polarization ellipse.

*Figure 1. Principle of Fraunhofer dark-line detection for fluorescence extraction. $\lambda_{in}$ represents the wavelength of the dark line of Fraunhofer absorption; $E_{in}$ and $E_{out}$ are, respectively, the absorption dark line and downward solar irradiance reaching the surface of the vegetation near the unaffected reference band on the left side of the absorption dark line; and $L_{in}$ and $L_{out}$ are the measured radiances of the band.*

*Figure 2. Flowchart of the chlorophyll-fluorescence inversion process (Huang et al. 2013).*
studying the polarized reflected light of natural features, the V vector is generally ignored (Talmage and Curran 1986). The radiation reaching the sensor through the polarizer can therefore be expressed by

\[ I(\alpha) = (I + Q \cos 2\alpha + U \sin 2\alpha)/2. \] (1)

where \( \alpha \) is the polarization angle, that is, the angle between the polarizer transmission direction and the reference direction. Other polarization components, such as the maximum and minimum radiation through the polarizer, can be obtained by

\[ L_{\text{max}}(\lambda) = \left[ I(\lambda) + \sqrt{Q^2(\lambda) + U^2(\lambda)} \right]/2 \] (2)

\[ L_{\text{min}}(\lambda) = \left[ I(\lambda) - \sqrt{Q^2(\lambda) + U^2(\lambda)} \right]/2. \] (3)

where \( L_{\text{max}} \) and \( L_{\text{min}} \) are the maximum and minimum radiation through the polarizer, respectively, which can be expressed by the Stokes vectors and the total radiation through the polarizer.

The polarization component \( L_{\text{pol}} \) can be expressed as a function of \( L_{\text{max}} \) and \( L_{\text{min}} \), or in terms of the Stokes vectors:

\[ L_{\text{pol}}(\lambda) = L_{\text{max}}(\lambda) - L_{\text{min}}(\lambda) = \sqrt{Q^2(\lambda) + U^2(\lambda)}. \] (4)

The unpolarized component \( L_{\text{unpol}} \) is derived from the radiation passing through the polarizer and the Stokes vectors; it can also be described as a function of \( L_{\text{pol}} \) and \( F \):

\[ L_{\text{unpol}}(\lambda) = I(\lambda) - \sqrt{Q^2(\lambda) + U^2(\lambda)} = L_{\text{pol}}(\lambda) + F(\lambda). \] (5)

The degree of polarization (DOP) can be measured as follows, where \( L_{\text{unpol}} \) represents the total radiation of vegetation reaching the sensor:

\[ \text{DOP} = \frac{L_{\text{pol}}}{L_{\text{unpol}}} = \frac{L_{\text{max}} - L_{\text{min}}}{L_{\text{max}} + L_{\text{min}}}. \] (6)

Using the sFIELD method (Wu and Zhao 2005), \( F \) can be separated from \( L_{\text{unpol}} \):

\[ F = \left( \frac{I_{\text{out}}/f_{\text{out}}}{I_{\text{in}}/f_{\text{in}}} \right)_{\text{unpol}} - \left( \frac{I_{\text{in}}/f_{\text{in}}}{I_{\text{out}}/f_{\text{out}}} \right)_{\text{unpol}}. \] (7)

where \( I_{\text{in}}/f_{\text{in}} \) and \( I_{\text{out}}/f_{\text{out}} \) represent the radiance values of the standard reference panel; \( f_{\text{out}} \) and \( f_{\text{in}} \) represent the reference reflectance factors, and \( L_{\text{in}} \) and \( L_{\text{out}} \) are the radiation values of vegetation in the two oxygen absorption bands—respectively, 761 and 756 nm for \( O_2 \) and 688 and 684 nm for \( O_3 \).

In this article, this model is used to verify the results. The SCOPE model has been widely used in recent years to interpret solar-induced chlorophyll fluorescence and investigate its links with photosynthesis links at different temporal and spatial scales. It is a nonvertical model that integrates the radiation transfer of solar radiation and the radiation emitted by the vegetation (thermal and chlorophyll fluorescence) with the energy balance (Guan et al. 2016). It can calculate directional canopy-reflected radiation, emitted thermal radiation, and solar-induced fluorescence signals as well as energy, water, and CO₂ fluxes (Van der Tol et al. 2009; X. Liu et al. 2019). This model needs some environmental parameters inputted to simulate fluorescence. The leaf composition, leaf area index, leaf inclination, and \( V_{\text{max}} \) (maximum carboxylation capacity at 25°C) are important factors affecting the solar-induced fluorescence simulation (Hu et al. 2018). The evaluation of the model supports the use of the SCOPE model as a powerful tool for simulating solar-induced chlorophyll fluorescence, and can provide an important reference for its future application.

**Measurement**

Experimental measurements were performed outdoors with a multi-angle measuring device, ASD spectrometer, and polarizer. The measuring device consists of three parts (Han et al. 2020): the base, a driven arm, and a circular ring. The base was placed on the 1.2-m-diameter circular ring and could achieve 360° rotation. The 1.2-m-long driven arm was then fixed on the base. The driven arm tilted up to 90° from the zenith and measures 0°–90° in the zenith direction. The view azimuth angle could also be adjusted from 0°–360° by turning the base. The driven arm and the circular ring were controlled to change the viewing zenith angles (0°–60°) and azimuth angles (0°–360°). For measurement, the samples were placed on the object stage. The spectral measurements were performed using an ASD FieldSpec 3 spectroradiometer, with a spectral range of 350–2500 nm. In the viewing direction, a bare fiber-optic cable was fixed in a metal tube and attached to the frame. The field-of-view angle of the spectrometer bare fiber was 25°; by extending the metal tube, the angle can be adjusted from approximately 5° to 25°. During the canopy polarization measurement, a polarizer was added to the front end of the detector fiber. Sunny and cloudless weather conditions were chosen for the experiment. In order to avoid the influence of the incident angle of the sun during the observation process, radiation was measured in both forward and backward directions when the solar incidence angle was 40.3°.

**Data**

Radiance data of the standard reference panel and Epipremnum aureum canopy were measured at different view azimuth and zenith angles (0°–360° and 0°–60°, respectively) by passing a polarizer that could be rotated to change the polarization direction (0° or 90°). Five spectral values were obtained simultaneously during each measurement. Any abnormal values were removed, and the average value was taken as the final spectral value. The acquired data were processed using ViewSpec software. The corresponding reflectance value of the Spectralon Diffuse Reflectance Standard—i.e., the bidirectional reflectance factor (BRF)—was obtained by calculation.

**Results and Discussion**

**Polarization Characteristics of Standard Reference Material**

In Figure 3a and 3b, DOP and \( L_{\text{pol}} \) are close to 0. In Figure 3a, the measured radiance spectra of the standard reference material in the two polarization directions (0° and 90°) agree closely with the calculated maximum and minimum radiance spectra \( L_{\text{max}} \) and \( L_{\text{min}} \) that pass through the polarizer, and they are basically coincident. This indicates that although the polarization direction is changed by rotating the polarizer, the reflected radiation intensity of the standard reference material received by the sensor is approximately the same. The total radiance calculated by Stokes vectors is approximately equal to the sum of \( L_{\text{max}} \) and \( L_{\text{min}} \). According to the polarization characteristics of light, when only unpolarized light passes through a polarizer, the radiation intensity in any direction is half the total reflected radiation intensity. The DOP is very small over this spectral range (close to 0). Figure 3b shows the polarization characteristics of the standard reference material at a different view zenith angle (60°). Here, DOP and \( L_{\text{pol}} \) are also very small and almost close to 0. The \( L_{\text{max}} \) and \( L_{\text{min}} \) values are close to the radiance values at 90° and 0°, respectively. Thus, despite the different view zenith angle, the polarization characteristics of the standard reference material remain consistent.
Polarization Characteristics of the Epipremnum aureum Canopy

The total radiance (calculated by Stokes vectors) and BRF of the Epipremnum aureum canopy indicates the typical spectral characteristics of green vegetation (Figure 4). The peak of total radiance is 0.65, and $L_{max}$ is only slightly larger than $L_{min}$. The total radiation intensity of the optical signal is theoretically equal to the reflected radiance under the same observation condition (Wu et al. 2013). In the BRF results, a reflection “bulge” caused by chlorophyll fluorescence is observed near 761 nm. The value of $L_{pol}$ is very small within the range of 400–900 nm and does not change with wavelength, which indicates that $L_{vol}$ dominates the apparent radiation.

The DOP reflects the proportion of polarized radiation in the total radiation. It is larger from approximately 400 to 500 nm and then starts to decrease close to 500 nm; then it begins to increase from 550 to 688 nm before decreasing again at approximately 688 nm. The maximum value of DOP is 0.3. Contrary to BRF, in areas with large DOP values the filling effect of polarized radiation on the absorption band cannot be ignored. Thus, in separating chlorophyll fluorescence from the reflected radiation, the influence of polarized radiation cannot be ignored. In Figure 5, which shows the same results at a view zenith angle of 60°, the BRF changes significantly at approximately 700–750 nm, with a maximum value of 0.6 at approximately 760 nm. The reflection “bulge” is still observed at 761 nm. Similarly, DOP begins to decrease at 688 nm, and its maximum value is 0.35.

Inversion of Solar-Induced Chlorophyll Fluorescence

Chlorophyll Fluorescence at O$_2$A (761 nm)

After the inversion calculation, we obtain chlorophyll fluorescence values of 0.0030–0.0035 and 0.0030–0.0040 at O$_2$A, respectively, for view azimuth angles of 0° and 180° (Figure 6). As the view zenith angle changes from 0° to 60°, the solar-induced chlorophyll fluorescence gradually increases. The ChlF at a view azimuth angle of 180° is slightly larger than that at 0°, reaching maximum values of 0.0039 and 0.0036, respectively. The ChlF values obtained from the inversion are consistent with those suggested by previous studies (Meroni et al. 2009; Huang et al. 2013).
Chlorophyll Fluorescence Inversion at O₂-B (688 nm)
The chlorophyll fluorescence values for O₂-B obtained from inversion are 0.0035–0.0045 and 0.0035–0.0040, respectively, for view azimuth angles of 0° and 180° (Figure 7). The ChlF at a view azimuth angle of 0° is slightly larger than that at 180°, with maximum values of 0.0040 and 0.0042, respectively, consistent with previous literature (X. Liu and Liu 2018). The ChlF generally increases in both forward and backward directions as the view zenith angle changes from 0° to 60°. Due to the influence of the canopy structure, the solar-induced fluorescence of the canopy is anisotropic, and changes in the view zenith angle will also affect the solar-induced fluorescence.

Verification
At the canopy level, the spectral resolution of the modeled spectrum is 1 nm, the reflection range is 400–2500 nm, and the fluorescence is 640–850 nm. In this study, the SCOPE model is used to simulate the fluorescence of the two bands O₂-A and O₂-B, as shown in Figure 8. The fluorescence values simulated by the SCOPE model are in the same range as those extracted from I₀ using the sFLD method. The simulated and calculated fluorescence values also exhibit the same trend. This shows the reliability of our method of chlorophyll-fluorescence inversion using polarization technology. The inversion accuracy is higher for O₂-B than for O₂-A: compared with O₂-B, the calculated and simulated values for the O₂-A

Figure 6. Chlorophyll fluorescence F at O₂-A (761 nm) for view azimuth angles of 0° and 180° and view zenith angles of 0°, 15°, 30°, 45°, and 60°.

Figure 7. Chlorophyll fluorescence F at O₂-B (688 nm) for view azimuth angles of 0° and 180° and view zenith angles of 0°, 15°, 30°, 45°, and 60°.

Figure 8. Calculated and simulated fluorescence values at a view azimuth angle of 0° and view zenith angles of −60° to 60° for (a) O₂-A (761 nm) and (b) O₂-B (688 nm).
band show larger deviation. This may be due to the unreasonable unbiased assumption of the solar incident radiation and the standard reference material [Hu et al. 2018]. In Figure 9, the high root-mean-square error for 761 nm indicates that the simulated and calculated values have large deviation, which shows that the chlorophyll fluorescence values in this band are not very good in the SCOPE model. Compared with 761 nm, the chlorophyll fluorescence values for 688 nm are better simulated by the SCOPE model. The smaller the root-mean-square error, the smaller the deviation between the calculated and the simulated value, and the better the simulation effect.

**Figure 9.** The root-mean-square error of calculated and simulated values in O$_2$-A (761 nm) and O$_2$-B (688 nm).

**Conclusions**

In this study, we separated polarized light L$_{pol}$ from unpolarized light L$_{unpol}$ using polarization technology, then employed the sFLD method to successfully extract chlorophyll fluorescence from L$_{unpol}$ at the absorption bands of O$_2$-B (688 nm) and O$_2$-A (761 nm). We obtained the following conclusions: a comparison with fluorescence values simulated by the SCOPE model indicated that polarization measurement can effectively separate volume scattering L$_{pol}$, polarized light L$_{unpol}$, and fluorescence emitted by chlorophyll (ChlF) from the apparent reflection radiance of a vegetation canopy. Decomposition of the vegetation-canopy apparent radiation provides the foundation for inverting ChlF. In both forward and backward directions, the calculated fluorescence gradually decreased from $-60^\circ$ to $0^\circ$, reaching a minimum at $0^\circ$, then gradually increased from $0^\circ$ to $60^\circ$. At the five view zenith angles selected in this study ($0^\circ$, $15^\circ$, $30^\circ$, $45^\circ$, and $60^\circ$), the chlorophyll fluorescence calculated from the inversion method was larger at O$_2$-B than at O$_2$-A. Both ChlF and L$_{pol}$ are diffuse scattered light, and L$_{unpol}$ accounts for a large proportion of the canopy apparent radiation, explaining most of the spectral characteristics of green vegetation. The L$_{pol}$ value was very small and less dependent on wavelength.

For vegetation research using remote-sensing technology, it is crucial to determine the polarized reflection characteristics of the vegetation canopy. In future research, we will analyze more types of vegetation canopy and consider the impact of additional environmental factors in order to more accurately invert chlorophyll fluorescence and improve the current reflectance-based vegetation model. Quantitative decomposition of radiance provides the potential to optimize vegetation remote-sensing inversion models. The relationship between the three components (L$_{pol}$, L$_{unpol}$, and ChlF) should be analyzed in more detail in order to develop a more accurate decomposition method. Unlike the previous inversion of chlorophyll fluorescence using the unpolarized method, this is a new exploration by our research group using polarization remote sensing in chlorophyll-fluorescence inversion. After the inversion, we used the SCOPE model to compare the accuracy and found that the polarization remote-sensing method of chlorophyll-fluorescence inversion is feasible. This is a new method that can be said to be an extension of the fluorescence-inversion method.

**Acknowledgments**

This research was supported by the Fundamental Research Funds for the Central Universities (2412018ZD012), the National Natural Science Foundation of China (41301364, 41630749), and the Foundation of the Education Department of Jilin Province (JJKH20211288KJ).

**References**


Han, Y., T. Hao and Z. Li. 2020. China patent no. ZL 2019 2 0881968.8. P.


Binbin He, Guoming Li, Liang Li, Miao Yang, Qiongyi Huang, and Zihan Guo. Ecological Functions and Human-activity Interferences Evaluation in Ecological Protection Redline for Urban Environment.

Fang Gao, Yihui Li, Wentao Li, Peng Zhang, Yuan An, Xing Zhong, and Yuwei Zhai. A High-Resolution Satellite DEM Filtering Method Assisted with Building Segmentation.

Xueyan Li, Yu Hou, Ruizeng Zhai, Junfeng Song, Xuehan Ma Shuzhao Hou, and Shuxiu Guo. Three-dimensional reconstruction of single input image based on point cloud.


Zezhong Zheng, Yi Ma, Yutang Ma, Mingcang Zhu, Ran Huang, Xueye Chen, Qingjun Peng, Yong He, Yufeng Lu, Guoqing Zhou, Zhigang Liu, Mujie Li. An Incremental Isomap Method for Hyperspectral Dimensionality Reduction and Classification.

Rongjun Qin, Xiao Ling, Ph.D., Xu Huang. A Unified Framework of Bundle Adjustment and Feature Matching for High-resolution Satellite Images.

Zezhong Zheng, Mujie Li, Mingcang Zhu, Yue He, Jun Xia, Xueye Chen, Qingjun Peng, Yong He, Xiang Zhang, and Pengshan Li. The spatio-temporal evolution of urban impervious surface for Chengdu, China.

Lei Zhang, Hongchao Liu, Xiaosong Li, and Xinyu Qian. Optimizing the segmentation of a high-resolution image by using a local scale parameter.


Qinghong Sheng, Rui Ren, Weilin Xu, Hui Xiao, Bo Wang, and Ran Hong. Spinor-based Attitude Determination with Star Sensor Considering Depth.

Bharath H Aitha and Prakash PS. Digital building height preparation from satellite stereo images.


Jiabao Li, M.D, Wei Han, Ruyi Feng, Lizhe Wang, and Fengpeng Li. Unsupervised representation high-resolution remote sensing image scene classification via contrastive learning convolutional neural network.

Longjie Ye, Wen Xiao, Yehua Sheng, Dong Su, Pengbo Wang, Shan Zhang, Na Zhao, Hui Chen, and Ka Zhang. Gaussian mixture model of ground filtering based on hierarchical curvature constraints for airborne LiDAR point clouds.


Chengming Ye, Hongfu Li, Ruiyong Wei, Lixuan Wang, Tianbo Sui, Wansen Bai, and Saied Pirasteh. Twice Adaptive Intensity Threshold Method for Uneven LiDAR Data to Extract Road Markings.

Majid Rahimzadegan, Arash Davari, and Ali Sayyadi Estimating regional soil moisture with synergistic use of AMSR2 and MODIS images.

Guoqing Zhou, Man Yuan, Xiaozhu Li, Hongjun Sha, Jiasheng Xu, Bo Song, and Feng Wang. Optimal regularization method based on L-curve for solution of the RPM parameters.


Quality Assessment of Heterogeneous Training Data Sets for Classification of Urban Area with Landsat Imagery

Neema Nicodemus Lyimo, Fang Luo, Qimin Cheng, and Hao Peng

Abstract
Quality assessment of training samples collected from heterogeneous sources has received little attention in the existing literature. Inspired by Euclidean spectral distance metrics, this article derives three quality measures for modeling uncertainty in spectral information of open-source heterogeneous training samples for classification with Landsat imagery. We prepared eight test case data sets from volunteered geographic information and open government data sources to assess the proposed measures. The data sets have significant variations in quality, quantity, and data type. A correlation analysis verifies that the proposed measures can successfully rank the quality of heterogeneous training data sets prior to the image classification task. In this era of big data, pre-classification quality assessment measures empower research scientists to select suitable data sets for classification tasks from available open data sources. Research findings prove the versatility of the Euclidean spectral distance function to develop quality metrics for assessing open-source training data sets with varying characteristics for urban area classification.

Introduction
Data quality is defined as a concept that includes data precision and accuracy to determine if the data are specific enough and the types or amount of errors they contain (Bielecka and Burek 2019). However, in this research, we are interested in a broader definition that embraces aspects of data relevance—data qualities that are often characterized as “fitness for use” (Stanislawski et al. 2014; Zhou et al. 2018; Shao et al. 2018). This article intends to assess the suitability of data extracted from multiple open-source platforms as training sets to classify or retrieve from remotely sensed images.

The increasing availability of crowdsourced data and other free open data sources brings new opportunities for geospatial applications (Deren et al. 2014, 2019; Yin et al. 2015; Shao et al. 2020). Despite its success, every source has its related challenges. For example, OpenStreetMap (OSM) data bring up new issues to consider, including variations of contribution patterns among regions caused by the “digital divide” between developing and developed countries (Goodchild 2007). As of October 2020, the OSM database in Europe was 22.1 GB but only 4.0 GB for the whole continent of Africa (http://download.geofabrik.de). Another aspect is its non-exhaustive nature. Users are more likely to contribute more to one specific place than to others due to, for example, familiarity and pride of place (Forget et al. 2018). Finally, there is the issue of quality, which is questionable concerning the contributors’ trustworthiness.

Open Government Data (OGD) is another initiative that has been picked up on worldwide, including in countries in developing regions (Vetrò et al. 2016). A study of seven African countries (Ghana, Sierra Leone, Tunisia, Morocco, South Africa, Kenya, and Tanzania) showed that by 2017, OGD Web portals had about 1500 data sets in total that were up to date and freely accessible online (Afful-Dadzie and Afful-Dadzie 2017; Lyimo et al. 2020). Unlike OSM, government data are authoritative and usually assumed to be of better quality (Fogliaroni et al. 2018). OGD data have countrywide coverage and are less affected by the drawbacks related to OSM data discussed in the previous paragraph. The downside of OGD is that publicly available data are usually the product of a derived data model (Shao and Li 2011; Lyimo et al. 2020). Hence, volunteered geographic information (such as OSM) and OGD have different modeling schemes and quality characteristics.

When presented with several free and open data sets, selecting the best data set to serve as the training set data becomes important. The selection of sample data points is an essential part of the supervised classification of remotely sensed imagery. The training data set’s quality is the key to the accuracy of classification results because inappropriate training samples are the primary source of classification errors (Pal and Mather 2006; Radoux et al. 2014; Shao et al. 2014). A study by Foody and Arora (1997) demonstrated that the choice of training samples significantly affects the classification results more than does changing the classifier model.

While most research has focused on quality measures for remotely sensed images, little work has focused on quality metrics for training data sets. Ge et al. (2008) proposed using rough set theory to analyze sample quality reliability for image classification problems. However, challenges related to selecting a discretization method for the decision table affect this method’s adaptation in a broader context. Another related work used open data Portuguese land cover Map (Cos) to...
generate training samples for random forest classifiers (Viana 2019). This study explored the k-means clustering technique to select the most representative training samples. These works assessed the quality of training data for image classification tasks from a single type of source or similar data sets. There is a research gap for assessing training data extracted from multiple heterogeneous data sets, particularly open-source–based training data sets.

Open-source data sets vary in many ways, such as in quality, quantity, and modeling schemes, to name a few. Assessment of training sets collected from different sources requires a robust procedure to assess the quality of samples and to optimize them for the classification task. In pixel classification, spectral analysis is the foundation of quality assessment. Various distance functions have been proposed for image quality evaluations. To identify a suitable distance function for accurately processing remotely sensed images, Deborah et al. (2015) compared existing distance functions, such as Manhattan, Chebyshev, and Euclidean distance functions; the spectral angle mapper; and the Levenshtein distance. Based on their results, they concluded that Euclidean spectral distance (ESD) is the most appropriate distance function. A study by Forget et al. (2018) applied the ESD function to assess the quality of training samples extracted from a single type of open source, namely, OSM, but the study did not discuss the efficiency of its evaluations. This article extends this technique to define the fitness of open-source training samples from volunteered geographic information and OGD. The study relies on the basic ESD principles to model different aspects of data quality in various contexts to effectively measure the quality of training data sets for image classification tasks. The proposed quality measures are evaluated on eight different data sets from two case study cities using Landsat 8 imagery.

**Case Studies and Data**

**Case Studies**

Dodoma and Arusha are two cities in Tanzania with different characteristics chosen as study sites (Table 1). Arusha is a major city with a temperate climate. It is a vibrant city that is considered an international diplomatic hub. In 2018, it was declared the capital of the East African Federation. On the other hand, tourism contributes a significant part to Arusha’s economy, making it Tanzania’s “safari capital” (Bigrube 2004).

In contrast to Arusha, Dodoma, the capital of Tanzania, is a growing city. It has been growing at a slower pace due to the delayed relocation of government activities. This city has a semi-arid climate (Shemsanga et al. 2016).

**Data**

**Satellite Imagery**

This study used Landsat 8 imagery from the US Geological Survey through the Earth Explorer website (https://earthexplorer.usgs.gov). The scenes were acquired as level 1 data products. Therefore, they are expected to be radiometrically calibrated and orthorectified (Forget et al. 2018; Shao et al. 2018, 2019; Twumasi et al. 2019). Table 2 shows the product identifiers and the acquisition dates of each scene. The latest set of cloud-free scenes were found for both study areas. Both scenes were acquired on the same date. For comparison purposes, it was found necessary to convert the DN values to surface reflectance values. The scene for each city was resized according to the area of interest (AOI) to reduce processing time.

**Built-Up Training and Validation Data**

OSM provides the most mature and reliable crowdsourced data in this region. The current literature shows that temporal accuracy, up-to-datedness, and lineage quality parameters of OSM in the Tanzania data sets are of higher quality in cities than in peripheral areas (Minghini et al. 2018). Therefore, we acquired OSM data for Dodoma and Arusha for August 2020 (Table 3). Data were downloaded via TurboPass. Building footprint layers were downloaded; other data objects were less represented and contained very little information, not sufficient to be used as training data.

**Table 1. Environmental and demographic characteristics of Dodoma and Arusha.**

<table>
<thead>
<tr>
<th>City</th>
<th>Climate</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dodoma</td>
<td>Semi-arid</td>
<td>Total population of 410,956 residents, according to the 2012 census (National Bureau of Statistics, Tanzania 2013).</td>
</tr>
</tbody>
</table>

**Table 2. Product identifiers and acquisition dates of each Landsat scene.**

<table>
<thead>
<tr>
<th>City</th>
<th>Landsat Product Identifier</th>
<th>Acquisition Date</th>
<th>Size of AOI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dodoma</td>
<td>LC08_L1TP_168064_20190929_20191017_01_T1</td>
<td>29 September 2019</td>
<td>51 x 47 km²</td>
</tr>
<tr>
<td>Arusha</td>
<td>LC08_L1TP_168062_20190929_20191017_01_T1</td>
<td>29 September 2019</td>
<td>14 x 13 km²</td>
</tr>
</tbody>
</table>

Tanzania’s OGD source is a rich collection that remains mostly untouched in geospatial applications (Lyimo et al. 2020). This collection contains spatial and nonspatial data. In this research, we were interested in spatial data that fall on the built-up area; hence, health facilities (HF), distribution points of water users (WDP), and school facilities (SF) data were selected (Table 3). The assumption behind the inclusion of water users’ distribution points is that they represent domestic users’ residential locations.

For pre-classification quality assessment, we used no more than 10% of OSM data to reduce processing time; the data were randomly selected across the entire region. However, pre-classification analysis for OGD included complete data sets since they are small in size.

Approximately 2900 polygons were digitized from very high spatial resolution imagery from Google Earth. The data were randomly collected throughout the entire area population to provide a standard measure for comparison of accuracy assessment. Forty percent of the samples (randomly selected) assessed the open data set fitness as training data for a supervised classification task. The remaining amount (60%) of the data set was used to evaluate the performance of built-up classification results.
Training and Validation Samples for Other Land Use/Land Cover Classes

This article's main goal is to assess the quality of open data as training samples; however, none of the sources listed in Table 3 had sufficient data to represent other land cover classes apart from the built-up class in both case study areas. Even if a study has a particular class of interest, conventional supervised classification requires that all categories that occur in the study area be included in the training stage to avoid substantial errors that may be difficult to detect even during accuracy assessment (Foody 2002; Foody et al. 2006). Therefore, we collected data from very high spatial resolution imagery from Google Earth for other land cover subclasses in each city, including water, farmland, bare land, vegetation fields, forest/trees, shrubs, and wetlands. After classification, these subclasses were combined into major classes, including water, bare land/low vegetation, and high vegetation.

Methodology

This methodology’s basic idea is to generate suitable quality assessment measures by considering the selected classifier’s requirements and by assessing variations in characteristics of training samples collected from different open data sources. Figure 1 provides an overview of the proposed methodology for quality assessment and validation.

Selection of Classifier

Several classification algorithms exist. We selected the maximum likelihood classifier (MLC), a simple, common classification algorithm that fits our the purpose, image characteristics, and training data of our analysis. MLC is a pixel-based classification approach that is based on Bayes’ theorem. It uses a discriminant function to assign pixels to the class with the highest probability (Ahmad and Quegan, 2012). It achieves that by calculating statistical distances based on the clusters’ means and covariance matrices (Ahmad and Quegan 2012; Stein and Tolpekin 2012). MLC is a supervised classification scheme that assumes that spectral classes are statistically characterized by their means and variances (Richards 1993). Statistical distances are probability values that measure spectral uncertainty, and a cell is assigned to the class (cluster) for which it has the lowest uncertainty.

Pre-Classification Training Data Quality Measures

Measuring Spectral Similarity

In pixel-based classification, individual image pixels are characterized by spectral information. Spectral classes represent surface characteristics, or land cover classes. The classification procedure considers a distance to the class’s mean as a key to deciding to which class to assign pixels; therefore, assessing spectral differences or similarities is important. We are evaluating the similarity of open-source training data sets.

Figure 1. An overview of the methodology.
with reference data. This section improves the ESD formula to determine similarity uncertainties for multiple heterogeneous training data sets. Let us consider k to be one of the heterogeneous training data sets in a given study area and R a reference data set for the study area under investigation. We measure each data set’s spectral signature S for the feature class object it represents in six nonthermal Landsat bands,

\[ S_k = (x_1, \ldots, x_a, \ldots, x_f), \]  

\[ S_b = (y_1, \ldots, y_a, \ldots, y_f), \]  

where \( S_k \) represents the spectral signature of the \( k^{th} \) data set with \( k = (1, 2, \ldots, m) \), \( S_b \) represents a spectral signature of a reference data set, and \( \bar{x}_a \) and \( \bar{y}_a \) are mean pixel values of the featured objects in the two data sets for band \( n \).

According to Forget et al. (2018), the ESD \( d \) between two featured objects \( x \) and \( y \) is given by

\[ d(x,y) = \sqrt{\sum_{n=1}^{N} (x_n - y_n)^2}. \]  

Since we are assessing each of the heterogeneous training data sets separately, Equation 3 can be rewritten to represent the cumulative ESD of data set \( k \):

\[ d_k(x,y) = \sqrt{\sum_{n=1}^{N} (x_n - y_n)^2}. \]  

We can then normalize the ESD \( d_k \) results between values \( a \) and \( b \) for each data set using the normalization Equation 5 to measure the basic similarity between the open-source training data sets and the reference data set. Therefore, we refer to this new measure \( d_k \) values as measures of similarity uncertainty (SimU). The smaller the SimU value, the higher the similarity of data set \( k \) to the reference data; hence, it is a higher-quality data set and vice versa:

\[ \text{SimU} = (b - a) \times \frac{d_k - \min(d_k)}{\max(d_k) - \min(d_k)} + a. \]  

**Distribution of Training Data Sets in the Feature Space**

In this section, we assess the ESD in a different context to determine the intensity of the distribution of feature points in the feature space. Ideally, it is considered that each data set’s pixel values will accumulate around certain areas in the feature space and form very dense clusters. The concentration of heterogeneous samples in the clusters will vary from one training data set to another. We assume that a data set whose feature points are very close to one another will have lower uncertainty than a data set whose feature points are far from one another (Figure 2). In other words, a data set whose feature points are very close will have higher quality than a data set whose feature points are far from one another.

This section modifies another Euclidean distance–based method proposed by a recent study of Zhang et al. (2019) for modeling uncertainties in remotely sensed images to facilitate uncertainty measurements of heterogeneous training samples in the feature space. The scholars applied the Euclidean distance formula to calculate the distance between feature points in the feature space, as shown in Figure 2. Our feature space is composed of \( N \) nonthermal spectral bands of Landsat imagery. We measure the intensity of distribution of the feature points of a given data set in the feature space as follows:

\[ \Phi_x = \frac{1}{m} \sum_{j=1}^{m} d_p, \]  

where \( \Phi_x \) represents distribution density of training data set points \( k \), \( m \) is the total number of pixels representing the feature points of a training data set \( k \), and \( d_p \) represents the ESD of the \( j \)th feature point of a training data set to the cluster’s reference center point \( p \). We calculate \( d_p \) using the following equation:

\[ d_p = \sqrt{\sum_{n=1}^{N} (f_n^a - f_n^p)^2}. \]  

where \( f_n^a \) represents the \( n \)th feature of the central reference feature point \( p \) in the cluster and \( f_n^p \) is the \( n \)th feature of the \( j \)th pixel of the training data point in the feature space; in this study, \( n \) corresponds to the six spectral nonthermal bands of a multispectral image, and the total number of dimension of feature space \( N \) is 6.

To obtain a measure for the assessment of feature space uncertainty (FSU) of the training data sets, we normalize the distribution density \( \Phi_x \) to values between \( a \) and \( b \) for each training data set \( k \) with Equation 8, and just like Zhang et al. (2019), we refer to this type of uncertainty as FSU:

\[ \text{FSU} = (b - a) \times \frac{\Phi_x - \min(\Phi_x)}{\max(\Phi_x) - \min(\Phi_x)} + a. \]  

**Integrating Spectral Similarity and Distribution Measures**

We have derived two variations of ESD-based measures from two different contexts or domains. To obtain a more comprehensive measurement model, we combine the two measures using a simple average formula. We refer to the resulting quality measure in Equation 9 as spectral uncertainty (SpU):

\[ \text{SpU} = \frac{\text{SimU} + \text{FSU}}{2}. \]  

Our analysis evaluates the effectiveness of these measures for ranking the quality of heterogeneous open-source training
data sets for classification tasks. The lower the uncertainty value, the higher the data set quality. The decision to include or exclude a particular data set as a training sample relies on the selected threshold. The assumption is that the lower the uncertainties, the lower the risk of using a low-quality training dataset and hence the higher the classification accuracy.

Classification System

The classification system was developed to represent major land use/land cover types based on the land surface’s heterogeneity, as shown in Table 4. The majority filtering process was applied to remove isolated unclassified pixels from the classification output. Majority analysis filtering is a standard smoothing procedure to reduce some salt-and-pepper noises (Su 2016).

Table 4. Land use/land cover classification system.

<table>
<thead>
<tr>
<th>Land Use/Land Cover</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Built-up</td>
<td>Residential/industrial/commercial areas where rooftops dominate</td>
</tr>
<tr>
<td>Bare land/low vegetation</td>
<td>Cleared land/farmland/bare land/areas with low vegetation growing</td>
</tr>
<tr>
<td>High vegetation</td>
<td>Areas covered with trees/shrubs/forest</td>
</tr>
<tr>
<td>Water</td>
<td>Water bodies, such as reservoirs, ponds, and rivers</td>
</tr>
</tbody>
</table>

Post-Classification Accuracy Measures

Even though pre-classification quality assessment measures provide a useful prediction of how the training samples will perform, the real effect of the given training data’s quality will be observed in the classification results (Ge et al. 2012). Therefore, we assess the correlation between pre-classification quality measures and post-classification quality measures. For comparison purposes, we evaluate the classification results of a given city with the same test set.

Here we consider accuracy measures, which allow us to effectively compare probabilities of either correct or incorrect classification for each result based on the training data set. The first measure is overall accuracy, referred to as a proportion of correctly classified pixels, given as (Foody 2002)

\[
P_o = \frac{\sum_{r=1}^{r-1} \sum_{j=1}^{j-1} p_{ij}}{\sum_{i=1}^{r} \sum_{j=1}^{j} p_{ij}}
\]

(10)

Another measure is the kappa coefficient. Kappa statistics are useful for evaluation and comparison classification results based on different data or methods (Shao et al. 2017). Let \( x_{ij} \) denote the element of the error matrix in row \( i \) and column \( j \), \( r \) denote the number of classes, and \( N \) denote the total sum of all elements of the error matrix. Then kappa coefficient \( k \) is computed as (Cohen 1960)

\[
k = \frac{N \sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r} \sum_{j=1}^{j} x_{ij}^2}{N^2 - \sum_{i=1}^{r} \sum_{j=1}^{j} x_{ij}^2}
\]

(11)

where \( x_{ii} = \sum_{j=1}^{j} x_{ij} \) and \( x_{ij} = \sum_{i=1}^{r} x_{ij} \) are the sums of all elements in row \( i \) and column \( i \), respectively.

Other measures include errors of omission and commission and user and producer errors. Omission errors refer to the number of pixels that were not included in interpreting the class results. In contrast, commission errors occur when samples have been wrongly classified as belonging to a particular class (Steelman and Czaplewski 1998; Stein and Tolpekin 2012; Sumari et al. 2020).

Analysis and Results

The study has derived quality measures based on ESD metrics to assess the fitness of four open-source training data sets for urban area classification in two cities. The open data are from two sources: OSM and OGD. We observe variations among the open-source data sets. For example, in OSM, we obtained building footprints (OSM BF) that contain 20 000 to 30 000 polygons. In OGD, we found three data sets that were related to built-up land use/land cover: SF, HF, and WDP. The amount of OGD data sets varied from hundreds to a few tenths. We also observed that different open data sources have variations in data types and formats. Data set types are of two categories: points and polygons. The data sets vary in quality, data size, and data type; considering all these variations, we would like to observe whether the proposed quality measures can successfully rank the quality of the training data sets for image classification.

Apart from built-up related data, we did not find enough freely available data for an accurate representation of other land use/land cover classes. Since we did not find such data in both study areas, we collected data from Google Earth for other land cover classes for each city. Training samples extracted from open sources were used for the classification of the built-up class. Therefore, quality assessments are carried out based on the variations of several open data sets used to train the built-up class.

Pre-Classification Quality Assessments

In the methodology, we derived three quality measures based on the ESD function: SimU, FSU, and a combination of the two, referred to as SpU. SimU measures the similarity between open-source training data sets and reference data using the mean of the ESD. FSU is a normalized measure of the spread or the concentration of the feature points of the data sets in the feature space. Finally, SpU takes advantage of the two measures’ varied capabilities by combining them to determine the overall spectral uncertainty of the training data sets.

Tables 5 and 6 show a summary of pre-classification quality assessments. A lower \( d_1 \) leads to a lower SimU value; hence, the data set is ranked as having a higher similarity in quality with the reference data set since it has lower similarity uncertainties and vice versa. On the other hand, when a data set cluster in the feature space is densely concentrated (\( \Phi_k \)), it leads to lower FSU; reflecting that the data set has good spectral coverage to represent a given training class. Values of 0.1 and 0.09 were used as scale values \( a \) and \( b \) in Equations 5 and 8.

The graphs in Figures 3 and 4 provide us with some visualizations of the variations in the data sets. We observe limited variations of spectral values for the data sets in Arusha compared to Dodoma. In Dodoma, the data are more similar in lower bands, but we notice a higher discrepancy between bands 5 and 7, except for OSM BF and reference data.

The WDP data set in Figure 3 has the largest difference from reference data in bands 6 and 7. Also, in Table 5, the WDP has a \( d_1 \) value of 0.099329; the difference is about seven times larger than OSM BF data. For clarity, we analyzed this section further in the two-dimensional feature space in Figure 5. The results show a significant shift of WDP data point for bands 6 and 7 in Dodoma.

The box chart type in Figure 5 enables us to picture and compare the distribution of the data sets by grouping them based on five fundamental values: minimum, first quartile, median, third quartile, and maximum. The chart’s box section is also referred to as the interquartile range (IQR); it represents 50% of the data values. The graph also shows the minimum and maximum spectral reflectance values in the data set via vertical statistical lines extending from the box.

OSM has the largest data sets in both cities, while HF data sets are the smallest, with the least having 41 data points.
Table 5. Summary of pre-classification quality assessments for Dodoma.

<table>
<thead>
<tr>
<th></th>
<th>OSM BF Data</th>
<th>SF Data</th>
<th>WDP Data</th>
<th>HF Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean spectral distance ($d_k$)</td>
<td>0.014 779</td>
<td>0.05 442 085</td>
<td>0.099 329</td>
<td>0.047 625</td>
</tr>
<tr>
<td>Similarity uncertainty (SimU)</td>
<td>0.1</td>
<td>0.475 088 969</td>
<td>0.9</td>
<td>0.410 788</td>
</tr>
<tr>
<td>Distribution of spectral values in the feature space ($\Phi_k$)</td>
<td>7.16E-07</td>
<td>0.000 355 692</td>
<td>0.000 242</td>
<td>0.001 162</td>
</tr>
<tr>
<td>Feature space uncertainty (FSU)</td>
<td>0.1</td>
<td>0.344 628 671</td>
<td>0.266 462</td>
<td>0.9</td>
</tr>
<tr>
<td>Spectral uncertainty (SpU)</td>
<td>0.1</td>
<td>0.40 965 882</td>
<td>0.583 231</td>
<td>0.655 394</td>
</tr>
</tbody>
</table>

Table 6. Summary of pre-classification quality assessments for Arusha.

<table>
<thead>
<tr>
<th></th>
<th>SF Data</th>
<th>OSM BF Data</th>
<th>HF Data</th>
<th>WDP Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean spectral distance ($d_k$)</td>
<td>0.0 111 979</td>
<td>0.031 139 503</td>
<td>0.012 942</td>
<td>0.028 260 712</td>
</tr>
<tr>
<td>Similarity uncertainty (SimU)</td>
<td>0.1</td>
<td>0.9</td>
<td>0.169 967</td>
<td>0.78 451 115</td>
</tr>
<tr>
<td>Distribution of spectral values in the feature space ($\Phi_k$)</td>
<td>4.57 057E-05</td>
<td>1.16 444E-06</td>
<td>0.000 168</td>
<td>5.72 079E-05</td>
</tr>
<tr>
<td>Feature Space uncertainty (FSU)</td>
<td>0.313 482 577</td>
<td>0.1</td>
<td>0.9</td>
<td>0.36 861 173</td>
</tr>
<tr>
<td>Spectral uncertainty (SpU)</td>
<td>0.206 741 288</td>
<td>0.5</td>
<td>0.534 983</td>
<td>0.57 656 144</td>
</tr>
</tbody>
</table>

Figure 3. Mean reflectance values for (a) Dodoma and (b) Arusha.

Figure 4. Distribution of spectral reflectance values for (a) Dodoma and (b) Arusha.
available only in Dodoma. However, we can observe that some data sets, such as SF, have comparable distribution coverage despite having a small size compared to OSM.

Classification Results
Classification accuracy measures were derived from the error matrix of classified images. The proposed post-classification measures include overall accuracy (OA), with all classes), overall kappa coefficient including all classes (OKC), built-up class accuracy (BA), kappa coefficient for built-up class (KC8), built-up unclassified pixels, a built-up area classified as bare land, a built-up area classified as an area with thick vegetation, and the errors of commission and omission for built-up class. Results of the OA, which include the OSM BF, OGD SF, OGD HF, and OGD WDP data sets, are 86.55%, 85.01%, 71.65%, and 59.44% in Dodoma and 85.86%, 86.52%, 84.42%, and 81.09% in Arusha, respectively (Figure 6).

Discussion
Role of the ESD for Modeling Quality Measures
The modeling quality measures depends on the assessment of spectral information in various contexts. Inspired by ESD metrics, we derived three quality measures for open-source training data sets: SimU, FSU, and SpU. Tables 7 and 8 present a visualization of a ranking correlation between pre-classification and post-classification quality measures.

A total of eight data sets were collected from OSM and OGD. The data sets are truly heterogeneous, with significant variations from one another in terms of quality, quantity, and data type. The evaluations are focused on the built-up class to determine the variations of different open data sets used for the training.

Tables 7 and 8 show a comparison between the proposed quality measures and overall built-up class accuracy. FSU ranks only five out of eight cases correctly, while SimU ranks six out of eight tested cases. When used alone, none of the two measures rank the quality of all the data sets successfully; each one is limited since it relies on a single perspective to assess the data. But when combined to produce a third metric, SpU, a more comprehensive quality assessment happens to rank all eight cases correctly. SpU performance proves the versatility of ESD for modeling different aspects of data to facilitate measuring the quality of training data sets for image classification purposes, especially those collected from diverse free sources.

Previous works have considered a single perspective for assessing data quality for training purposes; moreover, these works use data from single or similar sources (Forget et al. 2018; Viana 2019). Our results demonstrate that modeling the quality of training data from diverse open sources is more complex and takes a combination of different models for the successful quantification of quality. Spectral analysis is the key to examining data set quality for pixel-based classification. ESD has proven to be a powerful metric for modeling different aspects of quality measures on spectral similarity and FSU.

However, it should be noted that each data set should meet a minimum limit of data size for the selected classifier for a reliable SpU ranking. For example, MLC requires at least 10 p, where p is the number of bands used; for six bands, a data set with fewer than 60 pixels may have some inconsistencies, like in the case of HF data sets with 41 pixels for Dodoma. Its classification results have some discrepancies with kappa accuracy. However, some researchers in accuracy assessment have cautioned against using the kappa coefficient (Stehman...
and Czaplewski 1998); our future works will address this aspect further.

**Effect of Different Sizes of Training Sets**

In some cases, the increase in training data size has been shown to influence the decrease in FSU and even to minimize the overall SpU, leading to better classification results despite having a higher SimU. For example, Arusha OSM BF data with the largest SimU are ranked as the second-best data set by SpU. Compared with post-classification accuracy, it is also ranked as the second-best results in that city. It should be noted that OSM BF is the largest training data set in this city and so has smaller FSU. On the other hand, results show that when the data size is below a minimum requirement, significant increases occur in FSU, and overall performance drops drastically, such as in the case of classification results based on HF data points (37.13%). Despite the HF training class having a smaller SimU than SF data in Dodoma, the HF class size is almost half of the SF data set. Therefore, we can conclude that when all data sets are within the acceptable range of SimU, the amount of data size can positively influence the results. The derived measures can capture this aspect before classification begins and rank the data sets accordingly.

**Impact of Different Data Types**

The open-source training data sets comprise two main data types: points and polygons. For spectrally homogeneous land cover class, point-based training data sets have similar
performance to polygon-based training data (Chen and Stow 2002). For spectrally heterogeneous surface cover, polygons have the advantage of capturing a better range of available spectra information. However, pixels found along the borders of polygons usually contain mixed spectral values and can be detrimental to classification accuracy (Boudewyn et al. 2000). This phenomenon may impact the classification results of images with higher resolution than coarse resolution levels (Chen and Stow 2002). This factor’s influence cannot be noticed for small buildings whose area is less than or equal to a pixel for medium-resolution imagery like Landsat 8. However, it was interesting to discover that there are approximately 266 polygons whose area is larger than one or more pixels of Landsat 8 in Arusha. Despite the possible effect of mixed pixels in the boundaries, these polygons also have a higher chance of capturing pixels of higher purity levels. Hence, the polygons ensure more comprehensive coverage of the spectral range for a given training class.

This phenomenon is well modeled with FSU quality measures. Consider OSM BF and WDP results for Arusha. OSM BF is ranked with the highest level of uncertainty regarding SimU (0.9) and hence is considered a data set that is more dissimilar to the reference data set than WDP (SimU ≈ 0.78). However, this data set has the lowest FSU (0.1); a plausible explanation for this is the contribution of good spectral coverage caused by the polygon data type in this set. On the other hand, a reason for a lower SimU would be due to some mixed pixels of polygon boundaries, but it seems that this was not big enough to affect the overall quality of the data in SpU (0.5). Furthermore, point-based training data are more likely to have gaps in spectral information, causing a higher FSU and leading to an increase of unclassified pixels, as in the HF results in both case study areas.

**Conclusions**

Existing research has assessed training samples from a single or from similar open data sources. Such procedures may not be sufficient to examine the quality of heterogeneous data sets extracted from diverse sources. Eight different data sets from OSM and OGD were collected. We relied on ESD metrics to derive three quality measures for assessing the quality of the open-source training samples: SimU, FSU, and SpU. The lower the uncertainties, the higher the data quality. The study analyzes the relationship between pre-classification quality measures and post-classification accuracy measures.

Correlation analysis aimed to determine the robustness of the proposed quality measures against data sets of varying quality, data set size, and data types. A comparison with classification accuracy proves that SpU, which is the combination of FSU and SimU, can successfully rank the quality of training data sets despite their variations. These findings demonstrate the versatility of ESD for modeling different quality aspects of heterogeneous data sets based on spectral information in various contexts.

Other key findings include the following:

- Polygon data set type can help to reduce FSU because of continuous coverage of spectral information; however, at the same time, polygon boundaries can be the primary source of mixed pixels, which increases uncertainties in similarity measurements.
- Point data type can lead to higher FSU because of limited coverage of spectral information, especially for heterogeneous land surface covers, such as in urban areas.
- An increase in data set size has a positive influence and can even minimize the overall spectral uncertainty for data sets with low similarity uncertainties.
- However, data set size below the minimum requirements for a selected classifier can cause inconsistencies in the results.

This article provides a more comprehensive approach for the quality assessment of open-source training sets than existing works. With the increasing availability of open data, this method facilitates predetermination of the data set’s quality to empower researchers to choose suitable training sets for image classification. It also allows optimizing the data sets before the classification procedures take place.

This work has applied a simple average to combine two derived quality evaluation metrics to measure the spectral uncertainty of heterogeneous data sets in different contexts. Our future works will include experiments with different weights and combinations of parameters to further test and analyze the validity, applicability, and robustness of the proposed approach and to optimize it accordingly.

**Acknowledgments**

This work was supported by the National Key Research and Development Program of China (2018YFB0505401), the Research Project from the Ministry of Natural Resources of China under grant 4201-24010012, the National Natural Science Foundation of China under grants 41771452, 41771454, 41890820, and 41901340; the Natural Science Fund of Hubei Province in China under grant 2018CFA007; and National Major Project on High Resolution Earth Observation System (GFZX0403260306). The authors are sincerely grateful to the editors and anonymous reviewers for their valuable suggestions and comments that helped improve this article significantly.

**References**


Comparative Assessment of Target-Detection Algorithms for Urban Targets Using Hyperspectral Data

Shalini Gakhar and K.C. Tiwari

Abstract
Hyperspectral data present better opportunities to exploit the treasure of spectral and spatial content that lies within their spectral bands. Hyperspectral data are increasingly being considered for exploring levels of urbanization, due to their capability to capture the spectral variability that a modern urban landscape offers. Data and algorithms are two sides of a coin: while the data capture the variations, the algorithms provide suitable methods to extract relevant information. The literature reports a variety of algorithms for extraction of urban information from any given data, with varying accuracies. This article aims to explore the binary-classifier approach to target detection to extract certain features. Roads and roofs are the most common features present in any urban scene. These experiments were conducted on a subset of AVIRIS-NG hyperspectral data from the Udaipur region of India, with roads and roofs as targets. Four categories of target-detection algorithms are identified from a literature survey and our previous experience—distance measures, angle-based measures, information measures, and machine-learning measures—followed by performance evaluation. The article also presents a brief taxonomy of algorithms; explores methods such as the Mahalanobis angle, which has been reported to be effective for extraction of urban targets; and explores newer machine-learning algorithms to increase accuracy. This work is likely to aid in city planning, sustainable development, and various other governmental and nongovernmental efforts related to urbanization.

Introduction
Hyperspectral data are constituted of hundreds of narrow and contiguous bands which form a three-dimensional data structure known as data cube, holding a vast treasure of spectral and spatial information. The first two dimensions of this data cube correspond to the organization of pixels, specifying their location, and the third dimension relates to the associated spectrum, which is a function of wavelength in different bands. Hyperspectral imagery is extensively used in soil mapping (Schmid et al. 2016), wildlife and forest mitigation (Sims et al. 2013), water-body management (Keith et al. 2014), disaster prediction (Harb and Dell’Acqua 2017), mining (N. Li et al. 2018), urban land use classification (Man et al. 2015), and many other applications. In addition, hyperspectral data extend the scope of detection and identification for different military and civil targets. Targets can be characterized as objects of interest having distinct spatial and spectral attributes in contrast to the background. They may be of different sizes, confined within a set of pixels, a single pixel (full-pixel targets), or a fragment of a single pixel (subpixel targets).

A target-detection problem is regarded as a binary hypothesis according to which each pixel of the image is classified into a target or a nontarget background class. Target-detection algorithms can be broadly categorized into spectral-matching and anomaly-detection algorithms. In the former, the spectral signature along with additional details is already known by means of standard spectral libraries, ground truth, and laboratory analysis, or from the image itself. These algorithms include Spectral Angle Mapper (SAM), the spectral correlation measure, constrained energy minimization (CEM), and matched filter (MF). In the latter, no a priori knowledge about the target is available, and thus outliers—called anomalies—are detected by their unique behavior compared to background pixels. Orthogonal subspace projection, the Reed–Xiaoli detector, and the Uniform Target Detector are some examples of anomaly detectors. Apart from these techniques, which require comparison of spectra, independent component analysis is a new and evolving technique that aims to find components that are statistically as independent as possible. This does not depend whatsoever on the availability of spectra (Tiwari et al. 2011). Applications such as military surveillance (Yuen and Richard-son 2010), search-and-rescue operations (Eismann et al. 2009), and urban planning (Bassani et al. 2007) have necessitated further exploration of hyperspectral remote sensing.

Because urban areas are a blend of numerous material compositions, in the form of roads, roofs, parking lots, and other built-up surfaces, knowledge about them is required for the development of new infrastructure, analysis of aging monuments, population analysis, conservation, and mapping of resources. Conventional techniques for gathering these statistics are generally based on field surveys or manual analysis of aerial images, which are unable to keep up with the exponential development of urban cities (Heiden et al. 2007). Bridging this gap, hyperspectral remote sensing provides a less expensive and more time-efficient solution for automated identification of urban surfaces. Advancements using other sets of data, such as synthetic-aperture radar, and approaches like index-based methods are also used (Shao et al. 2014; Shao et al. 2016).

However, extraction of information pertaining to roads, roofs, and other urban materials is itself a challenge. For example, due to the effects of vibration and absorption, different materials can show similar characteristics, posing difficulty in distinguishing them. For instance, asphalt-based
roads and roofs are often confused with bare soil, and gray tile with composite shingle roofs (Herold et al. 2004). Also, data captured remotely show a higher level of intraclass variability due to varying object geometry, illumination effects, atmospheric interactions, spatial and spectral resolution, and many other parameters (Yadav et al. 2016; listed in Table 2). Therefore, mapping of urban targets is necessary for distinguishing between similar material types and land cover backgrounds. There is a need to reduce ambiguity in order to select the right algorithm to address target detection. Above all, spectral variability is an inherent characteristic observed among urban materials, where the spectral signature of the same target can vary not only with different geographical locations but also from sample to sample. However, the narrow hyperspectral bands are still capable of indicating salient absorption features which can be exploited in identifying the object class. And in target detection, an object can also be detected by distinguishing it from the background. Indistinctness may arise when the spectra of totally different materials appear to be the same, thereby increasing the false-alarm rate.

To date, most research appears to have been dedicated to natural targets such as vegetation and minerals; less contribution has been made to analyzing urban areas using remote sensing (Keith et al. 2014; Schmid et al. 2016). Due to identical material composition, specific urban targets may be spectrally similar, making it difficult to identify them solely on spectral grounds. The studies being conducted by the remote-sensing community have provided remarkable insight into the spectral properties of built-up materials, but a comprehensive analysis of the separability of different spectrally similar urban targets has yet to be addressed (Tiwari et al. 2013). Heterogeneity within the urban environment is another challenge, as the acquired signal is a result of multiple land surface targets, and it is therefore difficult to determine a suitable algorithm to map them. Target classes also present other unique problems in the case of urban targets. For example, most roads appear as linear features with narrow width, and roofs appear as clusters of small targets.

Hyperspectral data, due to their spectrally rich content, appear to be capable of providing solutions to these problems. A reasonable amount of research has already been done on existing hyperspectral urban data sets using a variety of approaches, including supervised/unsupervised classification, classification based on artificial neural networks (ANNs) or support vector machines, and object-based classification (Plaza et al. 2009). Support vector machines are also used in association with methods based on Markov random fields (Tarabalka et al. 2010) and as a semi-supervised learning method for hyperspectral images (Shao et al. 2014). The authors listed in Table 1 have used standard data sets such as Indian Pines (AVIRIS sensor), the University of Pavia (ROSIS sensor), and the Cuprite Mining District (HYDICE sensor), but publicly available data may not have significant signatures of regional materials, as materials used in the construction of urban surfaces differ from location to location. A multitude of surveys and reviews have been carried out (e.g., Shao and SrinivasaPerumal 2014), but implementation is lacking.

The objective of this article is to study an alternative approach for extraction of urban targets, treat it as a binary hypothesis, and make use of various target-detection algorithms to extract the targets. The novelty of the work lies in the idea of dealing with the problem of detecting urban targets by a target-detection approach rather than classification. Classification deals with multiple land cover classes at a time and generates a corresponding thematic map, whereas target detection searches for some specific objects, producing a binary image. Target detection is therefore well suited to detection of roads and roofs in urban areas. Before choosing to work on target detection, we tried out multiple supervised (parallelepiped, binary encoding, minimum distance) and unsupervised (K-means, ISODATA) classification methods, with overall accuracies ranging from 51% to 64%; however, to maintain a focus on target detection, we have not included those results in this article. A summary of various available target-detection algorithms is given in Table 1, along with their advantages and disadvantages. The present work sums up certain standard spectral-matching algorithms and performs an extensive comparative assessment for detecting urban targets (particularly roads and roofs). For this, all the spectral matching algorithms listed in Table 1 have been implemented, followed by accuracy assessment. The article also provides a review of conventional and modern target-detection techniques adopted in hyperspectral remote sensing in past years. The following section gives an insight into the theoretical background, then the data set used is described. After that comes a discussion of the methodology of the work, followed by the results, conclusion, and future scope.

Theoretical Background

Linear Spectral-Transformation Models

Dimensionality-reduction techniques enable us to transform hyperspectral data to a lower dimension and are applicable for data visualization. Being lossy in nature, these operations may affect detection and classification accuracy; but since hyperspectral data include a significant amount of spectral redundancy, some amount of dimensional reduction is appropriate. Spectral-transformation models enable the representation of vector data without any significant loss of information. The high dimensionality of hyperspectral data has two implications: the huge volume of data requires tremendous storage and processing resources, and the amount of data required for statistically oriented detection is greatly increased (Manolakis et al. 2016). The most popular spectral-transformation models include principal component analysis (Farrell and Mersereau 2005), vertex component analysis (Nascimento and Dias 2005), independent component analysis (Wang and Chang 2006), and linear discriminant analysis. Researchers have also tried sparse (Shao et al. 2014) and locality-preserving dimensionality reduction (W. Li et al. 2012) for hyperspectral image analysis preceding the detection process. Initially, an increase in the dimensionality of remote-sensing images favors classification accuracy, but that declines when the number of training samples is low. This problem is referred to as the Hughes phenomenon or Bellman’s curse of dimensionality (Hughes 1968). Dimensionality reduction is mentioned as part of preprocessing and the general approach for target detection; it is not the subject of study here. High-dimensional spaces have a huge volume, but data tend to occupy a very small subspace; in essence, high-dimensional spaces are mostly empty. As a consequence, high-dimensional data can be projected to a lower-dimensional subspace without significant loss of information, at least for classification applications. For target-detection applications, though, dimensionality reduction must be avoided or used with extreme care (Manolakis et al. 2016).

Analysis of Data-Collection Parameters for Target Detection

Detection and identification of targets involve gathering spectral signatures, which can uniquely determine the surface properties of the target or object under consideration. The spectral signature of a particular pixel is influenced by certain parameters such as illumination effects, height and background material, color, material composition, surface geometry of the object (slope, orientation, and texture), age of the material, and atmospheric interactions. Table 2 gives a brief introduction to factors that influence the spectral signature of a material.

Statistical Models for Target Detection

Statistical models for target detection are often used for their good performance, robustness, and ease of computation, and they provide a theoretical background for analyzing real-time
<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Approach</th>
<th>Advantages</th>
<th>Limitations</th>
<th>Data Sets Used</th>
</tr>
</thead>
</table>
| Euclidean distance (Robila 2005) | Provides a quantitative measure of the distance between the reference and unknown test spectra | • Computationally simple  
• Time efficient  
• Suitable for lower numbers of bands | • Values are positively defined and are not within a set interval  
• Distance increases with more bands | HYDICE  
Hyperion  
AVIRIS |
| Dot product (Stein and Scott 1994) | Provides the cosine angle between the target and test spectra | • Treats difference in peak intensities of reference and target spectra continuously to determine identical spectra | • More computation time required for evaluation  
• Relative intensities of neighboring peaks are not considered | NIST mass spectral database |
| Z-score (Parshakov et al. 2014) | Calculates the number of standard deviations a spectral signature is from a population or sample mean (spectral class) | • Takes into account the variation of pixel spectra within classes  
• Uses both mean and standard deviation, compared to other spectral-matching algorithms, which use only the mean | • Generates lower accuracy if not used with land cover types that have lower intraclass and higher interclass spectral variability | Landsat 5 TM image of an agricultural area |
| Mahalanobis distance (Imani 2019) | Based on a correlation which takes into account the mean and covariance of reference and test sample | • Efficient for unprocessed data | • Scene variability introduces substantial error in estimating parameters that describe heavy-tailed distributions  
• Making decisions by comparing whether the Mahalanobis distance is lower or higher than the fixed threshold is insufficient | AVIRIS |
| Spectral Angle Mapper (Falcone and Gomez 2005) | Supervised classification technique based on the computation of spectral-angle similarity between a reference source and the target spectra | • Advantageous in situations where target materials have different spectra | • Measures the angular direction of data points, not their magnitude  
• Relatively insensitive to illumination and albedo effects  
• Cannot distinguish negative from positive correlations, because only the absolute value is considered | NASA EO-1  
Hyperion, Level 1  
Radiance |
| Spectral correlation angle (Robila 2005) | The angle between correlation vectors of reference spectra and test spectra | • Allows detection of targets with negative correlation  
• Eliminates shading effect | • Evaluates the match based on the spectral shape, while ignoring the amplitude of the spectra | AVIRIS, Indian Pines test site |
| Spectral gradient angle (Angelopoulou et al. 1999) | The angle between the absolute values of gradient vectors of test and reference spectra | • Invariant to surface geometry, viewpoint, and illumination effects  
• Suitable for depicting the shape of the spectral curve | • Unable to handle noise in data | — |
| Mahalanobis angle (Manolakis et al. 2016) | The angle between the test and the reference vector after the whitening transformation | • Invariant under all nonsingular transformations | • The inverse of the correlation matrix cannot be calculated if the variables are highly correlated | — |
| Constrained energy minimization (Manolakis et al. 2003) | Constrains the energy of the desired spectral signature of the target by minimizing output energy from background | • Performs better with more bands | • The energy of the desired target level is lower than the energy of undesired pixels, resulting in false alarms | — |
| Matched filter (West et al. 2005) | A weighted inner product between the test pixel’s spectrum and the target spectrum | • Does not require knowledge of all the end members in the scene | • More false alarms | HYDICE data cubes from Forest Radiance I and Desert Radiance II |
| Adaptive cosine estimator (Truslow et al. 2014) | A measure of the cosine of the angle between the normalized test and target vectors | • Invariant to scaling of input data  
• Constant false-alarm rate | • Nonlinear in nature, which complicates statistical analysis | Real-time data set containing urban features |
| Spectral information divergence (Chang 2000) | A measure of divergence: if smaller, the target spectra belong to the reference class | • Efficient at preserving spectral properties compared to angular measures | • Pixels above the threshold value remain unclassified | AVIRIS  
HYDICE |
| Maximum-likelihood algorithm | Calculates the probability that a pixel belongs to a specific class | • Gives good results with fewer bands | • A band with no variance at all leads to a singularity problem  
• Assumes that the statistics for each class in each band are normally distributed | — |
| Artificial neural network (Pena et al. 2010) | Uses standard back-propagation algorithm applied to a set of input, hidden, and output layers | • Predicts results for unknown data sets | • Requires labeled data for training  
• Training of the network takes time | Indian Pines data |
| Extreme learning machine (Heras et al. 2014) | Learning algorithm for single-layer feed-forward neural network | • Higher accuracy than support vector machine and neural networks  
• Fast learning  
• Computationally scalable  
• Independent of tuning process | • Evaluation speed is low  
• Requires astronomically high number of hidden-layer neurons  
• Cannot encode more than one layer of abstraction | University of Pavia image (ROSIS)  
Indian Pines image (AVIRIS) |
These factors encompass selectable attributes, which can be varied to
increase the accuracy of unmixing, detection, identification, and clas-
sification of objects.

<table>
<thead>
<tr>
<th>Factor Type</th>
<th>Description</th>
<th>Parameters</th>
</tr>
</thead>
</table>
| Sensor-related       | Sensors record the amount of light reflected from the surface of the target according to numerous wavelength intervals. The efficacy of the response depends on following the necessary protocols of field spectroscopy. These factors can be taken care of at the time of preprocess-
ing, as few instruments allow configuration according to requirements. | Aperture size, focal length, focal-plane array size, spatial resolution, spectral resolution, signal-to-noise ratio, radiometric resolution, calibration, height and altitude of the instrument (Kerekes and Goldberg 2013) |
| Scene-related        | Targets are confined to very small number of pixels in a scene, and therefore the content of the spectra is influence by certain background features. | Material composition, color, scene complexity (number of distinct surface classes; Kerekes and Goldberg 2013), target condition (open/hid-
den/camouflaged/buried; Tiwari et al. 2013) |
| Atmosphere-related   | The behavior of electromagnetic radiation falling on a target is altered by absorption and scattering mechanisms of the atmosphere; simul-
taneous corrections are required to correct the distortions caused by atmospheric interactions. | Illumination, presence of haze, fog or clouds, temperature, moisture content (Kerekes and Goldberg 2013) |
| Spatial              | The spectral signature of a surface is highly dependent on spatial features such as location and size of the target. | Target location involves surface, subsurface, or air; target size includes a single pixel, a group of pixels, or subpixel (Tiwari et al. 2013). |
| Morphological        | Morphological features relate to the shape, boundaries, and convex hull of a specific target. | Point, linear (line), area (polygon) |
| Processing           | These factors encompass selectable attributes, which can be varied to increase the accuracy of unmixing, detection, identification, and classification of objects. | Threshold values, number of iterations, number of bands (Kerekes and Goldberg 2013) |

Target detection distinguishes objects of interest from back-
ground features. Its objective is to assign logical labels to every pixel vector in a hyperspectral image based on prior knowledge of reference spectra, where a value of 1 signifies the presence of a target and 0 represents a nontarget or back-
ground pixel. The targets of interest usually occupy a few pix-
els, and the algorithms enable a high possibility of effectively characterizing them by high-order statistics. The literature involves a collection of supervised, semi-supervised, and unsupervised approaches to target detection. In the context of supervised approaches, various distance, angle, information, and machine-learning measures are used to estimate the relation between the spectrum to be detected and the representa-
tives of the class (also shown in Table 1).

Distance Measures
The idea of spectral distance originates from the need for many hyperspectral applications to use measures to assess the similarity (or distance) between spectra, or between one spec-
trum and a group of spectra. Distance measures are frequently used in classification as well as target detection (Robila 2005). They are widely used to measure the similarity in shape of two spectral signatures. Keshava (2004) used distance matri-
ces for band selection in hyperspectral processing to identify materials. In the present study, the band-selection approach was applied to data for two spectrally similar classes, and the minimum-distance method succeeded in correctly discriminating pixels that were misclassified by other techniques, such as Euclidean minimum distance and least-square projection. The Mahalanobis distance is considered to be superior to the Euclidean distance, and is capable of exploiting correlation between hyperspectral data. It has recently been used by Imani (2019) as a difference-based target-detection technique, where an output is generated by calculating the difference between the distance of a test pixel from background and that of a test pixel from the target. Normalization of image values is neces-
sary, especially when distance classifiers are used. These meth-
ods aim to normalize each feature of the image in multiple ranges, thereby assigning equal weights to different features and reducing computational load (Naeini et al. 2017). The distance values computed for the entire data set are normalized between 0 and 1 by using the following formula, where $I$ is the lower bound, $u$ is the upper bound, $x_i$ is the vector at pixel $i$, and $x_n$ is the normalized matrix, ranging from 0 to 1:

$$\text{spdiff}\left(x_i, x_n\right) = \frac{x_i - I}{u - I}.$$  

Angle Measures
Robila (2005) has investigated the efficiency of various angular measures—spectral angle, spectral correlation angle (SCA), and spectral gradient angle (SGA)—when used with spectral screening of hyperspectral data, and concluded that the algo-
rithms implemented provide results that closely match those of processing on the full data and may therefore be considered as a substitute for the spectral angle. Angle measures have been used for target detection, with SAM being the simplest one. The algorithm determines the spectral similarity between two spectra by treating them as vectors in a space with dimen-
sionality equal to the number of bands and calculating the angle between them (Kruse et al. 1993). More recently, Imani (2019) has proposed two difference-based target-detection methods that contrast with standard detection algorithms. The first approach uses the Mahalanobis distance, whereas the second uses kernel-based SAM, utilizing the dif ference between target and background computed distances.

Information Measures
Cisz and Schott (2005) have evaluated the performance of hyperspectral target-detection algorithms for multiple targets and backgrounds, particularly CEM and the adaptive cosine estimator (ACE), in scenes at different altitudes (5000, 10 000, and 20 000 ft). Forest Radiance I data were collected with the HYDICE sensor, detecting grass, trees, and roads as targets. With the increase in availability of high-resolution data, it is important to have reliable and robust mechanisms to predict the performance of target-detection systems for real-time applications. For this purpose, a performance prediction model for the widely used MF and ACE detectors is presented using signal modeling (Truslow et al. 2014). Apart from these mea-
sures, spectral information divergence (SID) has also been used for spectral characterization (Chang 2000). Generally, statisti-
cal models use estimated parameter values which are optimal for observed data; hence, maximum-likelihood parameter
estimation is the basis of most of the target-detection algorithms, including MF, CEM, and ACE (Manolakis et al. 2003).

**Machine-Learning Measures**

More recently, machine-learning algorithms have shown remarkable results in the field, having exceptional capabilities to automatically learn the relationships between the data (here, spectra captured) and predict accordingly. These methods help in extracting useful information from raw data by generalizing the unknown facts. There are numerous algorithms that exploit the concepts of machine learning for analysis of hyperspectral data. Plaza et al. (2009) used neural network-based models for hyperspectral image separation. The focus of the work is on selecting small training sets as input to the network for characterizing mixed pixels. A new learning technique with a single-layer feed-forward network—known as an extreme learning machine (ELM)—has been developed by Huang et al. (2006). Heras et al. (2014) considered two ELM-based techniques integrating spectral and spatial image information. The first used a majority-vote approach to combine the results of a pixel-wise spectral classification, and the second introduced spatial information from a small spatial neighborhood after ELM classification. The standard spectral-matching algorithms used for analysis of target detection in this study are listed in Table 3 (see next page) with their mathematical formulas.

**Data-Set Description**

**AVIRIS-NG Data-Collection Campaign**

The hyperspectral data analyzed were acquired as part of the Airborne Visible and Infrared Imaging Spectrometer—Next Generation (AVIRIS-NG) data-collection campaign in February 2016. The area considered for investigation is a scene in Udaipur, Rajasthan, India, with a size of 1295×388 pixels, 425 bands in the range of 376–2500 nm, 5-nm spectral resolution, and 8.1-m spatial resolution. A Level 2 (derived geospatial variables at the same resolution and location as Level 1 source data) spatial subset of 394 lines and 385 samples, which constitutes the maximum urban area, was manually selected for the experiments.

**Ground-Data Collection**

The spectroradiometer used for ground-spectra collection for the same region spans 2151 channels with 1-nm spectral sampling. Figure 1 gives a representation of subset of size 394×385 pixels of the complete Udaipur image. The rectangle at the top middle represents the target roof, which is easily distinguishable from the background. The rectangle at the bottom left is an enlargement of highway road, and the rectangle at the bottom right is another type of roof. Table 4 provides a brief introduction to the image and ground-data collection parameters considered in this study.

**Methodology**

**Preprocessing**

As part of preprocessing the Level 2 AVIRIS-NG data with 425 bands, those bands that were severely affected by atmospheric gases and water vapor were removed. After removal of the bad bands, 387 remained for the implementation. According to the National Geographic Society, urban areas are very developed, meaning there is a density of human structures such as houses, commercial buildings, roads, bridges, and railways. A subset of existing infrastructure—roads and roofs—were chosen as urban targets to be detected.

**Target Detection**

The reference spectra for the implemented algorithms were derived in two ways: from ground data and from the image itself. In the first method, resampling of ground spectra (2151 channels) with respect to the image subset was performed, reducing them to 387 bands. As part of data collection in the field, multiple spectral signatures of urban targets were collected, including roads (bitumen, concrete), roofs (concrete, bitumen, corrugated galvanized iron sheets, sandstone), vehicles, and railway tracks. Out of these spectral signatures, those of concrete road and concrete roof were selected as references to distinguish on spectral grounds. The mean spectra of nine road samples and six roof samples were considered. In the second method, the mean spectra of both targets were computed by considering 100 pixels each from the hyperspectral image to serve as reference spectra in each algorithm. The geographic locations of the targets were also validated using Google Earth imagery, and ground spectral data collected. Various target-detection algorithms were implemented; their results are discussed later. Each algorithm produces a pair of outputs, comprising extracted road and roof pixels.

**Thresholding**

A multi-thresholding technique as an extension of Otsu’s method (1979) was used to apply a threshold to the resultant image based on minimizing the within-class variance in all implemented methods. The resultant image before thresholding can be represented with L levels ranging from 0 to 1. The number of pixels at level $i$ is denoted by $f_i$, and the total number of pixels sums to $N = f_0 + f_1 + \ldots + f_{L-1}$. For the resultant image, the probability of occurrence at the threshold $i$ is given by

$$p_i = \frac{f_i}{N}, \quad p_i \geq 0, \quad \sum_{i=1}^{K} p_i = 1 . \quad (2)$$

If an image is segmented into $K$ clusters ($C_0, C_1, \ldots, C_{K-1}$), then $K-1$ thresholds ($t_0, t_1, \ldots, t_{K-2}$) must be selected. The cumulative probability $w_k$ and mean gray level $\mu_k$ for each cluster $C_k$ are respectively given by:

$$w_k = \sum_{i \in C_k} p_i \quad \text{and} \quad \mu_k = \frac{\sum_{i \in C_k} i \cdot p_i}{w_k}, \quad k \in \{0,1,\ldots,K-1\} . \quad (3)$$

<table>
<thead>
<tr>
<th>Table 4. Description of image- and ground-data collection.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Parameter</strong></td>
</tr>
<tr>
<td><strong>Location</strong></td>
</tr>
<tr>
<td><strong>Date of acquisition</strong></td>
</tr>
<tr>
<td><strong>Spatial resolution</strong></td>
</tr>
<tr>
<td><strong>Spectral resolution</strong></td>
</tr>
<tr>
<td><strong>Channels</strong></td>
</tr>
<tr>
<td><strong>Level of data</strong></td>
</tr>
</tbody>
</table>

**Main features in the images**

Urban features such as roads, bridges, different types of roofs (concrete, bitumen, sandstone, corrugated galvanized iron sheets), vehicles, railway tracks.
Table 3. Mathematical formulas for target-detection algorithms.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Formula</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance measures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Euclidean distance</td>
<td>( e(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} )</td>
<td>( e(x, y) ): Euclidean distance ( x_i ): Test spectra ( y_i ): Reference spectra ( n ): Total number of pixels</td>
</tr>
<tr>
<td>Dot product</td>
<td>( x \cdot y = \sum_{i=1}^{n} x_i y_i + x_n y_n )</td>
<td>( x \cdot y ): Dot product ( y_i ): Test spectra ( y_i ): Reference spectra ( n ): Total number of pixels</td>
</tr>
<tr>
<td>Z-score</td>
<td>( ZSD = \frac{1}{n} \sum_{i=1}^{n} (t_i - t_{i\alpha})^2 )</td>
<td>( ZSD ): Z-score distance ( t_i ): Reflectance amplitude of n-band resampled reference spectrum at band ( b ) ( t_{i\alpha} ): Mean reference of n-band class spectrum at band ( b ) ( \alpha ): Class standard deviation at band ( b )</td>
</tr>
<tr>
<td>Mahalanobis distance</td>
<td>( Q = \frac{1}{n} \sum_{i=1}^{n} (s_i - \bar{s})(s_i - \bar{s})^T )</td>
<td>( Q ): Estimated covariance matrix computed with ( n ) training samples</td>
</tr>
<tr>
<td>Spectral Angle Mapper</td>
<td>( \text{SAM}(s, s_i) = \cos^{-1} \left( \frac{\sum_{i=1}^{n} s_i s_i}{\sqrt{\sum_{i=1}^{n} s_i^2 \sum_{i=1}^{n} s_i^2}} \right) )</td>
<td>( s, s_i ): Spectral vectors</td>
</tr>
<tr>
<td>Spectral correlation angle</td>
<td>( \text{SCA}(x, y) = \arccos \left( \frac{c(x, y) + 1}{2} \right) )</td>
<td>( x, y ): n-dimensional vectors ( x, \beta ): Expected values of vectors</td>
</tr>
<tr>
<td>Spectral gradient</td>
<td>( \text{SG}(x) = (x_1 - x_2, x_3 - x_2, x_4 - x_3, \ldots, x_n - x_{n-1}) )</td>
<td>( \text{SG} ): Spectral gradient ( x_i ): n-dimensional vectors ( \text{abs}(x) ): Vector whose components are absolute values of components of ( x )</td>
</tr>
<tr>
<td>Mahalanobis angle</td>
<td>( \cos \theta = \frac{(x - m)^T C^{-1} (x - m)}{\sqrt{(x - m)^T C^{-1} (x - m) (x - m)^T C^{-1} (x - m)}} )</td>
<td>( x, m ): Spectral vectors ( C ): Covariance matrix ( \theta ): Mahalanobis angle</td>
</tr>
<tr>
<td>Information measures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constrained energy minimization</td>
<td>( y_i = \mathbf{w}^T \mathbf{r} \quad t^R = 1 \quad w = \frac{R^T \mathbf{R}^{-1}}{\mathbf{R}^T \mathbf{R}^{-1}} )</td>
<td>( \mathbf{R} ): Autocorrelation matrix ( \mathbf{w} ): Coefficient vector</td>
</tr>
<tr>
<td>Matched filter</td>
<td>( x = x_0 - m_b \quad s = s_b - m_b )</td>
<td>( \text{MF} = \frac{x^T \mathbf{C}_s^{-1} s}{\sqrt{s^T \mathbf{C}_s^{-1} s}} )</td>
</tr>
<tr>
<td>Adaptive cosine estimator</td>
<td>( \text{ACE} = \frac{x^T \mathbf{C}_s^{-1} s}{\sqrt{s^T \mathbf{C}_s^{-1} s} \sqrt{x^T \mathbf{C}_s^{-1} x}} )</td>
<td>( x, s ): Spectral vectors ( \mathbf{C}_s ): Covariance matrix</td>
</tr>
<tr>
<td>Spectral information divergence</td>
<td>( I_r(x, r) = -\log(p(r)) \quad I_{t_i}(t_i) = -\log(q_i) )</td>
<td>( \text{SID} ): Divergence ( t_i, r_i ): Pixels with spectral signatures ( s_i ) and ( s_t ) ( p, q ): Probability functions generated by spectra ( s_i ) and ( s_t )</td>
</tr>
<tr>
<td>Maximum-likelihood algorithm</td>
<td>( g_i(x) = \ln p(Y) + \frac{1}{2} \ln \left[ 1 + \frac{1}{2} \left( x - m \right)^T \sum \left( x - m \right) \right] )</td>
<td>( x ): n-dimensional data ( \Sigma ): Covariance matrix</td>
</tr>
<tr>
<td>Machine-learning measures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Artificial neural network</td>
<td>( y = g_i = g_i \left( \sum \mathbf{w}_i x_i + \theta_i \right) )</td>
<td>( y ): Output ( x ): Input ( g_i ): Activation function ( \mathbf{w} ): Weight ( \theta ): Bias ( E ): Error ( t_i ): Ground truth</td>
</tr>
<tr>
<td>Extreme learning machine</td>
<td>( \sum_{i=1}^{L} \beta_i G(a_i, b_i x) \quad x \in \mathbb{R}^t )</td>
<td>( L ): Hidden nodes ( G(a_i, b_i x) ): Output function at the ( i )th hidden node ( a_i, b_i ): Hidden-node parameters ( \beta ): Weight vector ( g ): Activation function ( H ): Hidden-layer output matrix</td>
</tr>
</tbody>
</table>
Therefore, the mean intensity μₖ of the total image and the between-classes variance σ₂₀ are, respectively,

\[ \mu_k = \frac{1}{i=1} p_i = \frac{1}{i=0} \mu_i \nu_k \]  

\[ \sigma_0^2 = \frac{1}{i=1} \mu_k (\mu_k - \mu_i)^2 = \frac{1}{i=0} \nu_k \mu_k^2 - \mu_i^2. \]  

**Performance Evaluation**

A window of size 60×60 was selected containing road and roof as targets, as shown in Figure 2. Two binary images were created, with 1 = target and 0 = background, for the ground-truth image of road and roof. The spatial locations were validated by the field data and Google Earth coordinates. The known set of road pixels was calculated from ground-truth data to contain 33 pixels, and roof pixels totaled 42. The same region as for the reference window was extracted from the output subset of size 394×385 after thresholding. Using these values, the detection rate is evaluated as

\[ \text{Detection rate} = \frac{\text{pixels detected}}{\text{total pixels}} \times 100. \]

**Comparative Assessment**

A detailed comparative study was carried out to determine the most suitable algorithm for urban target detection. The comparison was made at two levels: between accuracies computed by ground-truth and image-based reference spectra, and among the various measures of target detection (distance, angle, information, and machine learning). A comprehensive bar graph comparing these measures was plotted, and a line diagram was made illustrating the threshold value range used for extraction of roads and roofs considering ground-based and image-based reference spectra.

**Implementation Background**

For implementation, various spectral-matching algorithms were used for detection and identification of the urban targets considered (roads and roofs). The software used to carry out the analyses was MATLAB 2018b and ENVI 5.0. The block diagram of the implementation is shown in Figure 3.

**Results and Discussion**

This article presents an alternative approach for extraction of urban targets: target detection, rather than the conventional classification method. The series of target-detection algorithms specified in Table 1 were implemented and a comparative assessment carried out.

The data set used for this study has already been discussed. A set of experiments were conducted based on the type of measure used: distance (Euclidean distance, dot product, Z-score, Mahalanobis distance), angle (SAM, SCA, CEM, MF, ACE, SID, maximum-likelihood estimation), and machine learning (ANN, ELM). The analysis was done in the following stages:

- **Stage 1**: Target detection using ground reference spectra.
- **Stage 2**: Target detection using spectra drawn from the scene.
- **Stage 3**: Analysis of thresholds for different algorithms.
- **Stage 4**: Performance evaluation.

**Stage 1: Using Ground Reference Spectra**

**Distance Measures**

The results obtained after implementation of various distance measures are shown in Figure 4 and Table 5. The distance between the spectra to be classified is calculated for all the representatives of the class, and the one with the minimum distance based on the threshold value is selected as the member of the class. The analysis with respect to the reference spectra taken for experiments was as follows:

- Euclidean distance. This measure gives reasonable accuracies for roads and roofs, but on visual analysis, the resultant images (Figure 4i, ii) do not highlight roads and roofs.
- Z-score. This measure is generally used for creating spectral libraries, but here it has been used to extract the urban targets. The resultant images do not appear to distinguish roads and roofs (Figure 4v, vi). But the measure performs slightly better than the Euclidean distance and the dot product.
- Mahalanobis distance. This measure achieves the best detection performance among distance measures. These observations indicate relatively poor performance for distance measures. These measures are all based on a simple linear modeling of spectra, which are, however, nonlinear in nature. This appears to be the reason for their poor performance. In addition, a reference spectrum obtained in the field is often influenced by factors such as sensor-related parameters, and atmospheric intervention, as noted in Table 2.

**Angle Measures**

Compared to distance measures, angle measures performed well (Figure 5, Table 6). Class pixel vectors at small angles
from reference spectra often belong to the same class (Harvey et al. 2002).

- **SAM.** This measure shows consistent performance but confuses a few pixels of roads and roofs at certain positions, as illustrated in Figure 5i and 5ii.
- **SCA.** On visual examination, SCA appears to detect patches of soil along with roads and roofs at the top right of Figure 5iii and 5iv.
- **SGA.** On the other hand, SGA shows no results contributing to extraction of roads and roofs (Figure 5v, vi). The detection rate also drops significantly, making it the least effective of these measures for detection.
- **Mahalanobis angle.** This measure outperformed the other angle measures, with an accuracy of 81.82% for road detection and 83.33% for roof detection. It has not been considered yet for extraction of targets using hyperspectral data, but these are promising results.

Angular measures appear to be efficient, as their detection rates are higher than those for distance measures. On experimentation, though, they took more computation time, with complex calculations. Loosely derived from distance measures, angle measures are combinations of dot products, differences, and so on, leading to better target detection. SAM achieved steady performance compared to the spectral correlation measure. Loss of spectral properties during calculation of the gradient might be a reason for the degraded performance of SGA. The Mahalanobis angle performs the best due to its consideration of covariance between neighboring bands.

### Information Measures

All the information measures were implemented with the intention of examining the influence of the image data set as input on the different techniques used for processing the data. The covariance matrix of the background is required to implement these algorithms; it is calculated using the entire image, assuming the target occupies a very small fraction of the complete image (Manolakis et al. 2001). In terms of the V-I-S model (vegetation, impervious surface, soil), vegetation and soil are considered background, whereas roads and roofs are considered targets (Ridd 1995). Figure 6 and Table 7 show the resultant images and accuracy, respectively.

- **CEM.** This measure is unable to detect any traces of road and roof pixels, according to visual analysis (Figure 6i, ii), and thus fails as a target detector. In road detection, it gives results that are highly confused with soil, whereas for roofs it generates an unclear image.
- **MF.** This measure’s output images (Figure 6iii, iv) are also highly jumbled with pixels of bare soil and produce unsatisfactory results.
- **ACE.** Visually, this algorithm shows results (Figure 6v, vi), but road pixels are wrongly detected as roof pixels.
- **SID.** The results of this measure are average; it is unable to differentiate between roads and roofs.

### Table 5. Accuracy assessment for distance measures using ground reference spectra.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Roads</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total Pixels</td>
<td>Pixels Detected</td>
<td>Detection Percentage</td>
<td>Total Pixels</td>
<td>Pixels Detected</td>
<td>Detection Percentage</td>
<td></td>
</tr>
<tr>
<td>Euclidean distance</td>
<td>33</td>
<td>17</td>
<td>51.51</td>
<td>42</td>
<td>22</td>
<td>52.38</td>
<td></td>
</tr>
<tr>
<td>Dot product</td>
<td>33</td>
<td>18</td>
<td>54.54</td>
<td>42</td>
<td>18</td>
<td>42.86</td>
<td></td>
</tr>
<tr>
<td>Z-score</td>
<td>33</td>
<td>19</td>
<td>57.57</td>
<td>42</td>
<td>22</td>
<td>52.38</td>
<td></td>
</tr>
<tr>
<td>Mahalanobis distance</td>
<td>33</td>
<td>23</td>
<td>69.69</td>
<td>42</td>
<td>30</td>
<td>71.43</td>
<td></td>
</tr>
</tbody>
</table>

### Table 6. Accuracy assessment for angle measures using ground reference spectra.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Roads</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total Pixels</td>
<td>Pixels Detected</td>
<td>Detection Percentage</td>
<td>Total Pixels</td>
<td>Pixels Detected</td>
<td>Detection Percentage</td>
<td></td>
</tr>
<tr>
<td>Spectral Angle Mapper</td>
<td>33</td>
<td>24</td>
<td>72.73</td>
<td>42</td>
<td>31</td>
<td>73.81</td>
<td></td>
</tr>
<tr>
<td>Spectral correlation angle</td>
<td>33</td>
<td>22</td>
<td>66.67</td>
<td>42</td>
<td>24</td>
<td>57.14</td>
<td></td>
</tr>
<tr>
<td>Spectral gradient angle</td>
<td>33</td>
<td>16</td>
<td>48.48</td>
<td>42</td>
<td>22</td>
<td>52.38</td>
<td></td>
</tr>
<tr>
<td>Mahalanobis angle</td>
<td>33</td>
<td>27</td>
<td>81.82</td>
<td>42</td>
<td>35</td>
<td>83.33</td>
<td></td>
</tr>
</tbody>
</table>

Figure 5. Detection of roads and roofs using angle measures and ground reference spectra.
• Maximum-likelihood algorithm. This measure is able to detect roofs, but does not perform well on extraction of road surfaces.

The information measures consider vegetation, impervious surfaces, and soil for computing the covariance matrix, but the scene may constitute of other urban features as well. Since the materials used for constructing roads and roofs may be the same, most of the algorithms appear to confuse roads with both bare soil and roofs.

Machine-Learning Measures

Machine-learning approaches show good generalization potential for detecting urban targets. They have exceptional capabilities to automatically learn the relationships between the data and to predict the desired results. In this work, ANN and ELM were used to analyze hyperspectral data. A training file of 2000 samples derived from ground data and the image was used for training the two networks. In the training file, 500 spectra each constitute the reflectance values of roads, roofs, vegetation, and soil. When the algorithm was implemented for a particular target (road or roof), all pixels belonging to the other categories were treated as background or nontarget. The training function used was back-propagation for ANN and a positive hard-limit transfer function for ELM. The hidden neurons varied from 20 to 50 for ANN and 950 to 1000 for ELM. Training parameters such as performance error, epochs, momentum, and learning rate, were tuned by trial and error in order to improve network accuracy. The images are shown in Figure 7, with the corresponding detection percentages in Table 8.

• ANN. Learning from the data provided leads ANN to produce good results, with an accuracy of 84.84% for roads and 88.09% for roofs (Figure 7i, ii).

• ELM: Similarly, detection of roads and roofs is prominent for ELM. A well-defined boundary of road and roof can be visualized in the output images (Figure 7iii, iv).

Despite the data complexity of hyperspectral imagery, machine-learning measures give the best results of all measures—with the added advantage that when predicting on an unknown data set, these algorithms attain higher detection rates.

The algorithms were implemented considering ground reflectance spectra as the reference. Due to a lack of knowledge about material compositions, sufficient target data were not available on the ground. The reference spectra obtained in the field were taken under different atmospheric conditions and are highly influenced by the background. Also, the test spectra to be compared for target detection were derived from hyperspectral imagery. Despite resampling the target spectra with respect to the image test spectra, a mismatch between the two persists. As already discussed, distance, angle, information, and machine-learning measures were implemented using a ground-based reference signature. Of the four, machine-learning algorithms, namely ANN and ELM, showed remarkable detection rates. Free of ANN’s drawback of

---

**Table 7. Accuracy assessment for information measures using ground reference spectra.**

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Roads</th>
<th>Roofs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constrained energy minimization</td>
<td>Total Pixels Detected 33, Detected 20, 60.61%</td>
<td>Total Pixels Detected 42, Detected 27, 64.28%</td>
</tr>
<tr>
<td>Matched filter</td>
<td>Total Pixels Detected 33, Detected 21, 63.64%</td>
<td>Total Pixels Detected 42, Detected 28, 66.67%</td>
</tr>
<tr>
<td>Adaptive cosine estimator</td>
<td>Total Pixels Detected 33, Detected 26, 78.79%</td>
<td>Total Pixels Detected 42, Detected 31, 73.81%</td>
</tr>
<tr>
<td>Spectral information divergence</td>
<td>Total Pixels Detected 33, Detected 19, 57.57%</td>
<td>Total Pixels Detected 42, Detected 30, 71.43%</td>
</tr>
<tr>
<td>Maximum likelihood</td>
<td>Total Pixels Detected 33, Detected 17, 51.51%</td>
<td>Total Pixels Detected 42, Detected 26, 61.90%</td>
</tr>
</tbody>
</table>

---

**Table 8. Accuracy assessment for machine-learning measures using ground reference spectra.**

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Roads</th>
<th>Roofs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Artificial neural network</td>
<td>Total Pixels Detected 33, Detected 28, 84.84%</td>
<td>Total Pixels Detected 42, Detected 37, 88.09%</td>
</tr>
<tr>
<td>Extreme learning machine</td>
<td>Total Pixels Detected 33, Detected 29, 87.87%</td>
<td>Total Pixels Detected 42, Detected 36, 85.71%</td>
</tr>
</tbody>
</table>
Stage 2: Using Image Reference Spectra

Distance Measures

The output images for roads and roofs are shown in Figure 8, and detection accuracies are given in Table 9.

- Euclidean distance. This measure detects road and roof surfaces with accuracies of 63.64% and 69.05%, respectively. In roof extraction, it highlights the bare soil area at the bottom right of the image (Figure 8i, ii).
- Dot product. This measure is unable to produce significant results for either road or roof extraction. In both cases, it produces similar visual results (Figure 8iii, iv).
- Z-score. This measure appears to detect roads with an accuracy of 66.67% but is not equally efficient at roof detection.
- Mahalanobis distance. This measure excels at determination of roof surfaces—even a small target is detected (Figure 8viii)—but in road detection it also detects surfaces covered with bare soil (Figure 8vii).

The reason for these performance results may be confusion between the similar materials used for construction of roads and roof—concrete, bitumen, and so on—leading to false alarms. Apart from this, an additional step of data normalization increased the computation time, as the computed values did not fall within a set interval. Consequently, the distance values increase with an increase in the number of bands (Keshava 2004).

Angle Measures

Figure 9 and Table 10 illustrate the results of SAM, SCA, SGA, and the Mahalanobis angle.

- SAM. According to visual examination (Figure 9i, ii), SAM is able to differentiate between road and roof surfaces.
- SCA. This measure partly recognizes the targets.
- SGA. This measure did not show good results even after contemplation of slope changes within a vector.
- Mahalanobis angle. This measure generated effective detection accuracy for both roads and roofs, but so far it has not been productively used for hyperspectral target-detection applications.

Despite having the added advantage of being insensitive to illumination changes, SAM was unable to perceive difference in intensities of spectral reflectance for broadly similar materials, and yielded similar spectral angles for road and roof detection. SCA performs well when the number of similar spectra is lower, but it considers negative correlations (Nidamanuri et al. 2018).

Table 9. Accuracy assessment for distance measures using image reference spectra.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Roads</th>
<th></th>
<th></th>
<th>Roofs</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total Pixels</td>
<td>Pixels Detected</td>
<td>Detection Percentage</td>
<td>Total Pixels</td>
<td>Pixels Detected</td>
<td>Detection Percentage</td>
</tr>
<tr>
<td>Euclidean distance</td>
<td>33</td>
<td>21</td>
<td>63.64%</td>
<td>42</td>
<td>29</td>
<td>69.05%</td>
</tr>
<tr>
<td>Dot product</td>
<td>33</td>
<td>23</td>
<td>69.70%</td>
<td>42</td>
<td>27</td>
<td>64.28%</td>
</tr>
<tr>
<td>Z-score</td>
<td>33</td>
<td>22</td>
<td>66.67%</td>
<td>42</td>
<td>26</td>
<td>61.90%</td>
</tr>
<tr>
<td>Mahalanobis distance</td>
<td>33</td>
<td>25</td>
<td>75.76%</td>
<td>42</td>
<td>33</td>
<td>78.57%</td>
</tr>
</tbody>
</table>

Table 10. Accuracy assessment for angle measures using image reference spectra.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Roads</th>
<th></th>
<th></th>
<th>Roofs</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total Pixels</td>
<td>Pixels Detected</td>
<td>Detection Percentage</td>
<td>Total Pixels</td>
<td>Pixels Detected</td>
<td>Detection Percentage</td>
</tr>
<tr>
<td>Spectral Angle Mapper</td>
<td>33</td>
<td>25</td>
<td>75.76%</td>
<td>42</td>
<td>31</td>
<td>73.81%</td>
</tr>
<tr>
<td>Spectral correlation angle</td>
<td>33</td>
<td>19</td>
<td>57.57%</td>
<td>42</td>
<td>20</td>
<td>47.62%</td>
</tr>
<tr>
<td>Spectral gradient angle</td>
<td>33</td>
<td>20</td>
<td>60.61%</td>
<td>42</td>
<td>27</td>
<td>64.28%</td>
</tr>
<tr>
<td>Mahalanobis angle</td>
<td>33</td>
<td>29</td>
<td>87.88%</td>
<td>42</td>
<td>36</td>
<td>85.71%</td>
</tr>
</tbody>
</table>
al. 2010). SGA was not able to distinguish between roads and roofs, perhaps because of highly redundant band information. The Mahalanobis angle achieved the best detection and was successful at extracting the urban targets considered.

**Information Measures**

The resulting images are shown in Figure 10, and the corresponding accuracies are given in Table 11.

- CEM. This measure detects roof pixels when it is supposed to be detecting roads, leading to much lower accuracy.
- ACE. This measure performs equally well for both roads and roofs, with respective accuracies of 81.82% and 83.33%.
- SID. This measure shows remarkable results for detection of roofs, with an accuracy of 78.57%, thereby preserving the spectral characteristics of a class.
- MF and maximum-likelihood algorithm. These measures are partially successful at detecting roads and roofs, detecting some areas of soil as target (Figure 10iii, iv, ix, x).

On visual examination, information measures produced satisfactory results and were able to extract roads and roofs as urban targets. ACE surpassed the detection accuracies of other algorithms in this category, whereas the maximum-likelihood algorithm was able to detect only traces of targets.

**Machine-Learning Measures**

The results for both algorithms are presented in Figure 11 and Table 12.

- ANN. This measure achieves remarkable accuracy: 87.87% for roads and 90.48% for roofs.
- ELM. This measure has been recently explored with hyperspectral data for its high accuracy (in the present experiment, 90.9% for roads and 83.33% for roofs) along with ease of implementation, less training time, ability with large data sets, and minimal human intervention.

Large amounts of data lead to better accuracy for machine-learning methods. The only challenge is the availability of labeled data. With known data, both the algorithms considered in this category performed well and were able to detect the targets. The range of values considered for feature vectors of roads and roofs overlap, making derivation of discriminative features difficult.

The measures in this stage make use of a reference signature derived from the scene and then compare it to the test spectra from the image itself. The detection percentage was higher than when the algorithms took field reference spectra into account. The reference signature in this case was tightly related to the imagery, and variability encountered was only in terms of the spectral range of values for a particular class. Also, the problem of insufficient spectral signatures did not arise in this case. Therefore, a large set of spectra combined to give an optimal reference spectrum. The common observation in both scenarios is the high performance of machine-learning algorithms, which are considered more suitable for these data.

**Stage 3: Analysis of Thresholds for Different Algorithms**

The results produced are binary images after thresholding, with black pixels indicating 0 (target absent) and white indicating 1 (target detected). Once the detections were made by all 15 target-detection algorithms, the detected pixels per image were compared to the ground-truth images. Two ground-truth images were generated, one each for roads and roofs.

![Figure 10. Detection of roads and roofs using information measures and image reference spectra.](image)

![Figure 11. Detection of roads and roofs using machine-learning measures and image reference spectra.](image)

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Roads</th>
<th>Roofs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Pixels Detected</td>
<td>Pixels Detected</td>
<td>Detection Percentage</td>
</tr>
<tr>
<td>Constrained energy minimization</td>
<td>33 24</td>
<td>72.73</td>
</tr>
<tr>
<td>Adaptive cosine estimator</td>
<td>33 27</td>
<td>81.82</td>
</tr>
<tr>
<td>Spectral information divergence</td>
<td>33 24</td>
<td>72.73</td>
</tr>
<tr>
<td>Matched filter</td>
<td>33 23</td>
<td>69.70</td>
</tr>
<tr>
<td>Maximum likelihood</td>
<td>33 15</td>
<td>45.45</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Roads</th>
<th>Roofs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Pixels Detected</td>
<td>Pixels Detected</td>
<td>Detection Percentage</td>
</tr>
<tr>
<td>Artificial neural network</td>
<td>33 29</td>
<td>87.87</td>
</tr>
<tr>
<td>Extreme learning machine</td>
<td>33 30</td>
<td>90.9</td>
</tr>
</tbody>
</table>

**Table 11. Accuracy assessment for information measures using image reference spectra.**

**Table 12. Accuracy assessment for machine-learning measures using image reference spectra.**
mapped with the locations of ground-spectra collection and validated with Google Earth coordinates.

A set of variable thresholds was generated by Otsu’s method (1979), as already explained; the minimum value was used, and the areas to the right of the threshold on the false-alarm and target histograms were measured. The resulting map was thresholded to distinguish likely targets and backgrounds, creating a binary image which was compared to the ground-truth image to evaluate the performance of the target detector. Table 13 shows the threshold values used for detecting roads and roofs using ground and in-scene reference spectra. Figures 12 and 13 present corresponding line charts.

Stage 4: Performance Evaluation

The results obtained by the target-detection algorithms were arranged in histograms. Figures 14 and 15 illustrate the effectiveness of machine-learning algorithms, which perform well for AVIRIS-NG hyperspectral data. The other observation is that when reference spectra are based on field data, the accuracies of all the algorithms decline by certain amount relative to when they are derived from the scene itself. This may be due to the influence of parameters such as background, moisture content, temperature, and many more during spectrum capture in the field.

Conclusion

This article presents an alternative approach for extracting urban targets using hyperspectral data. The problem of extraction of urban targets—roads and roofs—was successfully dealt with as a target-detection problem and studied in detail. A comprehensive taxonomy of target-detection methods using hyperspectral remote-sensing was presented. Several algorithms were also used that have been reported in the literature but not implemented for target detection, such as the Mahalanobis angle. Several target-detection algorithms were implemented using recently acquired hyperspectral AVIRIS-NG data from the Udaipur region of India.

The algorithms were compared in four categories: distance measures (Euclidean distance, dot product, Z-score, Mahalanobis distance) angle measures (SAM, SCA, SGA, Mahalanobis angle), information measures (Constrained energy minimization, Adaptive cosine estimator, Spectral information divergence, Matched filter, Maximum-likelihood algorithm), and machine-learning measures (Artificial neural network, Extreme learning machine).

Table 13. Threshold values considered for detection of road and roof targets.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Ground Reference Spectra</th>
<th>Image Reference Spectra</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Road</td>
<td>Roof</td>
</tr>
<tr>
<td>Distance measures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Euclidean distance</td>
<td>0.1686</td>
<td>0.0902</td>
</tr>
<tr>
<td>Dot product</td>
<td>0.1843</td>
<td>0.1569</td>
</tr>
<tr>
<td>Z-score</td>
<td>0.1882</td>
<td>0.1411</td>
</tr>
<tr>
<td>Mahalanobis distance</td>
<td>0.2353</td>
<td>0.1725</td>
</tr>
<tr>
<td>Angle measures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spectral Angle Mapper</td>
<td>0.1757</td>
<td>0.1608</td>
</tr>
<tr>
<td>Spectral correlation angle</td>
<td>0.2666</td>
<td>0.1568</td>
</tr>
<tr>
<td>Spectral gradient angle</td>
<td>0.149</td>
<td>0.1294</td>
</tr>
<tr>
<td>Mahalanobis angle</td>
<td>0.1372</td>
<td>0.1448</td>
</tr>
<tr>
<td>Information measures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constrained energy minimization</td>
<td>0.2039</td>
<td>0.1965</td>
</tr>
<tr>
<td>Adaptive cosine estimator</td>
<td>0.1255</td>
<td>0.0915</td>
</tr>
<tr>
<td>Spectral information divergence</td>
<td>0.0941</td>
<td>0.1019</td>
</tr>
<tr>
<td>Matched filter</td>
<td>0.2392</td>
<td>0.1921</td>
</tr>
<tr>
<td>Maximum-likelihood algorithm</td>
<td>0.1451</td>
<td>0.1243</td>
</tr>
<tr>
<td>Machine-learning measures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Artificial neural network</td>
<td>0.999</td>
<td>0.998</td>
</tr>
<tr>
<td>Extreme learning machine</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>
angle), information measures (CEM, MF, ACE, SID), maximum likelihood), and machine-learning measures (ANN, ELM).

Roads and roofs were treated as urban targets.

The work was implemented in different stages, using reference spectra acquired from the field in stage 1 and spectra drawn from the scene itself in stage 2. The experiments conducted for target detection show that roads and roofs can be effectively extracted from hyperspectral data using both the field reference spectra and the scene-drawn spectra. However, in all implementations, the reference signature drawn from the image produced higher detection rates, whereas the field reference spectra due to effects of background, illumination, and resolution yielded low detection rates. The reason for this difference could be the fact that down-sampling of ground data with respect to image data degrades the performance of the target detector. In addition, the spatial resolution of data used in these experiments was 8.1 m, which might be insufficient for detecting urban targets, as their scale is much smaller. A sharp delineating boundary between roads and roofs was therefore not seen in the resulting images.

Machine-learning methods appear to be more attractive, and show their tremendous potential to detect targets in less execution time and at higher accuracy rates. Their ability to learn from known data and predict for unknown data allowed them to outperform all other algorithms. Distance measures were less effective, and are insensitive to spectral properties of roads and roofs. Information measures led to partial detection of roads and roofs, as all the techniques depend upon calculation of the background covariance matrix. Mahalanobis-angle target detection performed well in comparison to other angle and distance measures but did not outperform machine-learning methods. Overall, machine-learning algorithms appeared uniformly to perform high; various target-detection algorithms in other categories produced mixed results; and some performed well for roads while others performed well for roofs.

Target detection depends hugely on the threshold value. Here, Otsu’s method (1979) for generating multiple thresholds was explored for generating binary output maps of target and background values. The threshold values varied by material composition, image data set, and algorithm used. Another important observation is that class distributions of the hyperspectral imagery are non-Gaussian, and thus a class may not be defined by a single component. Because of this, it is difficult to map urban targets using ground or image reference signatures. Apart from this, materials used for the construction of roads and roofs are similar, such as concrete, asphalt, or bitumen, which could be another reason for lower detection rates.

In summary, a target-detection approach appears to yield effective results in extracting roads and roofs using hyperspectral data.

Future Scope
Removal of redundant bands is needed to realize a better target-detection system. Dimensionality-reduction methods can be explored for decreasing the computation time of all algorithms. As observed, consideration of only spectral aspect leads to confusion in classes made up of similar materials; therefore, spatial attributes that exhibit strong dependencies in hyperspectral data should be considered for enhanced detection. In recent times, many techniques for selecting and extracting features have been developed which can be used in association with machine- and deep-learning measures for real-time processing of data. With the availability of hyperspectral data comes enormous information which may open new gateways in this field.

Acknowledgments
The work was in collaboration with the Defense Research and Development Organization and was sponsored by the Department of Science and Technology. The authors extend their gratitude for the support.

References


A Real-Time Photogrammetric System for Acquisition and Monitoring of Three-Dimensional Human Body Kinematics

Long Chen, Bo Wu, Yao Zhao, and Yuan Li

Abstract
Real-time acquisition and analysis of three-dimensional (3D) human body kinematics are essential in many applications. In this paper, we present a real-time photogrammetric system consisting of a stereo pair of red-green-blue (RGB) cameras. The system incorporates a multi-threaded and graphics processing unit (GPU)-accelerated solution for real-time extraction of 3D human kinematics. A deep learning approach is adopted to automatically extract two-dimensional (2D) human body features, which are then converted to 3D features based on photogrammetric processing, including dense image matching and triangulation. The multi-threading scheme and GPU-acceleration enable real-time acquisition and monitoring of 3D human body kinematics. Experimental analysis verified that the system processing rate reached ~18 frames per second. The effective detection distance reached 15 m, with a geometric accuracy of better than 1% of the distance within a range of 12 m. The real-time measurement accuracy for human body kinematics ranged from 0.8% to 7.5%. The results suggest that the proposed system is capable of real-time acquisition and monitoring of 3D human kinematics with favorable performance, showing great potential for various applications.

Introduction
Real-time capture and response for human locomotion at a large scale is of great importance for various applications, such as monitoring actions of patients in physical rehabilitation (Karunarathne et al. 2014), enhancing safe conditions of workers in industrial robotics (See et al. 2015), analyzing the movements of athletes (Gholami et al. 2019), and human-computer interaction in virtual reality (Jaimes and Sebe 2007). Because such applications benefit from accurate extraction and analysis of 3D human body kinematics, real-time photogrammetric systems capable of these types of measurements have been extensively researched in recent years.

With the advances of computer processing capabilities, human pose recognition has been shifted from single image-based (Agarwal and Triggs 2006; Shotton et al. 2011) to image sequences (Zhou et al. 2016), and further evolved to dynamic human pose recognition from video sequences (Wang et al. 2019). Nevertheless, the complexity of these algorithms prohibited the real-time processing of human post recognition. The recent development of smart cameras (Carraro et al. 2016) and red-green-blue-depth (RGB-D) sensors (Tang et al. 2020; Wu et al. 2019; Tang et al. 2016), which can directly capture 3D information in a given scenario, has enabled cost-efficient estimation and tracking of 3D body posture in real time. However, the sensors are limited by the workable distance, field-of-view (FOV), and reliability. In contrast, ordinary RGB cameras are capable of capturing large-scale scenarios with a large FOV. Previous studies using RGB cameras only recognized human posture in 2D (Shotton et al. 2011; Jalal and Kim 2014), resulting in the loss of vital information in the depth dimension. Multi-view stereo techniques (Seitz et al. 2018) enabled retrieving 3D information from 2D images using photogrammetric approaches, and 3D human body kinematics can be subsequently extracted from the retrieved 3D information. However, 3D reconstruction of large-scale scenes using dense image matching (Haala 2013) is time-consuming, especially for systems without hardware acceleration. The complexity of dense image matching algorithms requires balancing the efficiency and quality of the matching results, which impedes the possibility of real-time processing.

The advent of the central processing unit (CPU) with multi-threading capabilities and the development of the GPU-acceleration technologies make real-time computations possible. This paper presents a cost-effective photogrammetric system consisting of a stereo pair of RGB cameras. The system utilizes multi-threading and GPU-acceleration techniques as well as deep learning to extract and measure 3D human body features at a large scale in real-time, which enable further analysis of 3D human kinematics, such as step length, moving speed, arm angle, knee angle, etc., from the video sequence.

The remainder of this paper is organized into four sections. Section “Related Works” consists of reviews of related works, and the section “System Development” provides detailed descriptions of the developed system. The experimental evaluations are presented in the section “System Implementation and Evaluation”, and the section “Conclusions and Discussion” consists of concluding remarks and suggestions for future work.

Related Works

2D Human Body Feature Extraction
A good algorithm for human body feature extraction can improve the efficiency and accuracy of human body tracking.
and monitoring system. Early studies of human-computer interaction (Jaimes and Sebe 2007) introduced variable computer vision algorithms, such as body, gesture, gaze, and facial expression recognition algorithms, into several crossroad research areas, including psychology, artificial intelligence, and many others. These algorithms turned 2D human body feature detection into an intensive research field and applied to areas such as the facial feature point-recognition method (Xiong and De la Torre 2013) and single- or multiple-person posture recognition (Zhou et al. 2013). Xiong and De la Torre (2013) applied the facial feature recognition method and supervised descent method (SDM) to an image sequence. The SDM was able to recognize the facial features in the image sequence with favorable accuracy. Zhou et al. (2013) presented a gesture tracking and recognition algorithm, which allowed near real-time processing. Their experiments indicated that the running frame rate reached five frames per second (fps). Recently, the accelerated advancement of computer technology and the evolution of multithreading-capable CPUs have led to the popularity of deep learning approaches (Ranjan et al. 2017), such as mask regional-based convolutional neural networks (r-CNNs) (Abdulla 2017), OpenPose (Cao et al. 2018), and regional multi-person pose estimation (RMPE) (Fang et al. 2017). Ranjan et al. (2017) presented an algorithm called HyperFace, which allowed simultaneous face detection and posture estimation using deep CNN. However, HyperFace required three seconds to process one image, which limited its potential for real-time human feature extraction. OpenPose and RMPE have also made it possible to extract and extract 2D features of human postures in real time. Fang et al. (2017) used the benchmark Max Planck Institute for Informatics (MPI) human pose data set (Andriluka et al. 2014) and Microsoft Common Objects in Context (MSCOCO) data sets (Veit et al. 2016) to compare popular leading-edge human pose estimators based on the mean average precision (mAP) score. Table 1 provides an overview of these popular human pose estimators. According to Fang et al. (2017), deep learning-based object-detection and pose-evaluation algorithms accurately obtained the 2D key points of human posture. Among the assessed algorithms, RMPE was the most reliable and accurate multi-person pose estimator with an overall mAP of 80+ and a processing rate of 20+ fps. The OpenPose algorithm had an mAP of almost 70+ but only achieved approximately 10+ fps running on the same platform (Cao et al. 2018). Due to high process efficiency and accuracy, deep learning approaches are particularly suitable for real-time 2D human posture evaluation and feature extraction.

### 3D Human Posture Feature Extraction

In recent years, the rapid developments of computer hardware (D’Apuzzo 2002) and affordable RGB-D cameras (Zimmermann et al. 2018) have expanded the study of human posture evaluation and feature extraction from 2D to 3D space. D’Apuzzo (2002) proposed a method using photogrammetry to recover 3D human body features from synchronized video sequences captured from multiple cameras at different locations and dynamically tracked their trajectories. The creation of a 3D human kinematic descriptor (Zanfir et al. 2013) moved the study of 3D human body posture recognition and feature extraction from part-based posture retrieval methods (Zimmermann and Brox 2017; Sridhar et al. 2013) to whole-body human pose estimation (Srivastav et al. 2018). Even though these studies have brought human feature detection into 3D space, they were in general time-consuming and had not been implemented for real-time 3D human feature detection.

RGB-D cameras offer 3D information in a direct way and have been used for extracting 3D human posture and features in recent years. For example, Carraro et al. (2018) and Huang and Nguyen (2019) used the OpenPose to RGB-D camera to obtain 3D human feature points by integrating the 2D features extracted by deep learning with the depth information measured by the depth sensor. Srivastav et al. (2018) used an RGB-D camera to obtain 3D human body key points for indoor posture-estimation and tracking. However, the use of RGB-D cameras is limited by their short measurement ranges and narrow FOVs (Haggag et al. 2013).

This paper presents a real-time photogrammetric system consisting of a stereo pair of RGB cameras. The system incorporates a novel multi-threading strategy and CPU acceleration as well as an advanced deep learning method to extract and measure 3D human kinematics in real-time. The main contributions of the presented work are as follows:

1. In order to achieve real-time processing, the system adopts four threads, responsible for 3D scene reconstruction, human feature extraction, kinematic information calculation, and result visualization, respectively. Each thread works independently without waiting for other threads to complete their tasks, and thus the processing latencies of the procedures are reduced. In addition, for the thread of 3D scene reconstruction, which is a bottleneck problem and requires the most computations, a GPU-accelerated strategy is used to achieve real-time efficiency of 3D reconstruction.

2. To avoid complicated 3D computations, we develop a strategy that combines a 2D human body skeleton extraction algorithm with the projection relationships between 2D and 3D spaces. A mature deep learning method is adopted to ensure the reliability and efficiency of 2D human feature extraction, and then the 2D results are converted into 3D space based on the projection relationships, enabling 3D human body kinematic analysis.

### System Development

#### Multi-Threading Design

The real-time photogrammetric system had four threads. Each thread performed as an individual model that handled different tasks, as shown in Figure 1. Thread 1 loaded the stereo RGB images with timestamps and known orientation parameters, and then delivered images to semi-global matching (SGM) (Hirschmuller 2007) to generate a disparity map. A 3D map was retrieved by triangulation based on the disparities and orientation parameters of the camera. A GPU-acceleration solution was also used in the first thread to speed up the 3D scene reconstruction processing rate. Thread 2 first extracted 2D

---

Table 1. Comparison of 2D human detection and tracking algorithms (all values are mAP scores).

<table>
<thead>
<tr>
<th>Head</th>
<th>Shoulder</th>
<th>Elbow</th>
<th>Wrist</th>
<th>Hip</th>
<th>Knee</th>
<th>Ankle</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fang et al. (2017) (RMPE)</td>
<td>88.4</td>
<td>86.5</td>
<td>78.6</td>
<td>70.4</td>
<td>74.4</td>
<td>73.0</td>
<td>65.8</td>
</tr>
<tr>
<td>Iqbal and Gall (2016)</td>
<td>58.4</td>
<td>53.9</td>
<td>44.5</td>
<td>35.0</td>
<td>42.2</td>
<td>36.7</td>
<td>31.1</td>
</tr>
<tr>
<td>Insafutdinov et al. (2016) (DeeperCut)</td>
<td>78.4</td>
<td>72.5</td>
<td>60.2</td>
<td>51.0</td>
<td>57.2</td>
<td>52.0</td>
<td>45.4</td>
</tr>
<tr>
<td>Levinkov et al. (2017)</td>
<td>89.8</td>
<td>85.2</td>
<td>71.8</td>
<td>59.6</td>
<td>71.1</td>
<td>63.0</td>
<td>53.5</td>
</tr>
<tr>
<td>Insafutdinov et al. (2017) (ArtTrack)</td>
<td>88.8</td>
<td>87.0</td>
<td>75.9</td>
<td>64.9</td>
<td>74.2</td>
<td>68.8</td>
<td>60.5</td>
</tr>
<tr>
<td>Cao et al. (2018) (OpenPose)</td>
<td>91.2</td>
<td>87.6</td>
<td>77.7</td>
<td>66.8</td>
<td>75.4</td>
<td>68.9</td>
<td>61.7</td>
</tr>
<tr>
<td>Newell et al. (2017)</td>
<td>92.1</td>
<td>89.3</td>
<td>78.9</td>
<td>69.8</td>
<td>76.2</td>
<td>71.6</td>
<td>64.7</td>
</tr>
</tbody>
</table>
human body skeletons from the left-view images using RMPE and extended the 2D skeletons to 3D body features based on the 3D map array produced by thread 1. Thread 3 computed the following human kinematic parameters: moving velocity, step length, and joint motion angles. These parameters were based on 3D body features. The products of each thread were stored in the same queue for data exchange, and the results were loaded into thread 4 from the queue for the system visualization in real-time.

**3D Scene Reconstruction from Stereo RGB Images with GPU Acceleration**

This section describes the algorithms used in thread 1 for dense image matching and the triangulation process with GPU acceleration. The GPU-accelerated procedure is shown in Figure 2. First, stereo RGB images with known interior and exterior orientation parameters were loaded from the stereo camera and stored in the host (CPU). The device (GPU) then copied the stereo RGB images from the host and split them into left-view and right-view images. The dense image matching algorithm SGM and triangulation process were performed on the GPU with an accelerated solution for reconstructing the 3D information in real-time. Simultaneously, the left-view image and disparity map obtained from SGM were stitched together as the background image prepared for the visualization in thread 4.

**GPU-Accelerated SGM for Disparity Estimation**

A GPU-accelerated SGM method was applied for the real-time stereo estimation of the disparity map. Figure 2 shows the generation of a consistent disparity map. Two cost items, matching cost and smoothed cost, were computed in the GPU device. The matching cost measures the probability that two pixels on the left- and right-view images correspond to the same point in the object space. Features were first extracted from the left- and right-view images, and a similarity comparison was used to generate a local-matching cost and potential disparity for each pixel. A center-symmetric census transform (CSCT) (Hernandez-Juarez et al. 2016) configured with a fixed-sized (e.g., 9 × 7) window was used to extract the features by moving the window on the left- and right-view images, respectively. The extracted features were presented as bit-vectors. The similarity between two corresponding pixels in the left- and right-view images was defined as the Hamming distance between their CSCT bit-vector features.

The smoothed cost was introduced to deal with nonunique or incorrect correspondences of similarity, resulting in an inaccurate estimation of disparity. In SGM, the smoothed cost was computed by considering the similarity between neighboring points and disparities along paths across the image. The global solution was approximated as one-dimensional minimization problems along these paths. For each path direction, SGM aggregated a cost that considered the cost of neighboring points and disparities. After the disparities of all pixels were estimated, a 3 × 3 median filter (Brownrigg 1984) was applied to remove outliers.

GPUs are massively parallel devices containing tens of streaming multiprocessors (SMs), and vector computation operations are highly utilized and pipelined in SMs to optimize the computational efficiency. The compute unified device architecture (CUDA) programming model (Nvidia 2019) allows for defining a massive number of threads deployed in SMs of the same program code. The SGM was coded using a two-level identifier in CUDA to specialize in each thread for disparities estimation. The code in this research was deployed following the method of Hernandez-Juarez et al. (2016).

**Triangulation for Generation of 3D Map**

Triangulation was used to generate a 3D map (point cloud) from the disparity derived from the stereo camera. Figure 3 shows the geometry of the stereo camera system. C₁ and C₂ are the perspective centers of the left and right cameras, respectively, and I₁ and I₂ are the respective image planes. f₁ and f₂ are the focal lengths of the left and right cameras. They are the same for the camera system used in this research.

Figure 3 illustrates the geometric relationships between the object point P and the stereo cameras C₁ and C₂, and the colinear relationship amongst the object point, image point.
and camera perspective center (e.g., \( P, p_1, \) and \( C_1 \); or \( P, p_2, \) and \( C_2 \)). Based on the geometric relationships, the 3D coordinates of the point \( P \) can be calculated using the following equation (Kaehler and Bradski 2016):

\[
\begin{bmatrix}
X_p \\
Y_p \\
Z_p
\end{bmatrix}
= \begin{bmatrix}
u_1 - c_{1x} \\
v_1 - c_{1y} \\
f_1
\end{bmatrix} \frac{b}{d}
\]

(1)

where \((X_p, Y_p, Z_p)\) are the coordinates of the point \( P \) in object space. \((u_1, v_1)\) are the image coordinates of \( P \) in the left image. \((C_{1x}, C_{1y})\) are the coordinates of the principal point in the left image. \( f_1 \) is the focal length of the left camera. \( b \) is the baseline between the left and right cameras, and \( d \) is the disparity as denoted by \( u_1 - u_2 \). It should be noted that Equation 1 is used here instead of the complex collinearity equations for the purpose of more efficient calculation. Equation 1 is based on the epipolar geometry, which can be derived from the collinearity equations (Fraser 1997; Gruen and Beyer 2001), assuming a fixed relationship between the left and right cameras. In the actual experiments, the camera system used has been calibrated by the manufacturer already. The focal length of the camera, the position of the principal point, and lens distortions are provided. The images have been rectified based on the lens distortion parameters. A fundamental matrix defining the relative orientation of the left and right cameras is also provided, which allows the determination of epipolar geometry between the left and right cameras.

Based on Equation 1, each pixel in the disparity map can be transferred to a 3D point in the object space. The calculated 3D points can be used to generate a 3D map, and the RGB information of the 3D points can be obtained from the corresponding 2D coordinates for visualization purposes. Figure 4 shows an experimental result of 3D map visualization.

**Extraction of 3D Human Body Features**

Thread 2 handled the extraction of 3D human body features. The system took advantage of the mature 2D body skeleton extraction algorithm, RMPE (AlphaPose) (Fang et al. 2017), and then extended the 2D body skeleton into 3D body features based on the projection relationship between the image space and object space. RMPE is an open-source CNN-based multi-person pose estimator used in conventional pictorial structure models for posture estimation. RMPE has been evaluated on two standard multi-person data sets with large occlusion cases: MPII (Andriluka et al. 2014) and MSCOCO 2016 Keypoints Challenge data set (Veit et al. 2016). MPII data set contains more than 28,000 training samples for single person pose estimation, while the MSCOCO data set consists of 105,698 training and around 80,000 testing human instances. The results of RMPE on MPII data sets indicated that it achieved an average accuracy of 72 mAP on identifying difficult joints such as wrists, elbows, ankles, and knees. The results of RMPE on MSCOCO data sets also proved that RMPE achieved state-of-the-art performance compared with other popular detectors (Fang et al. 2017). Since RMPE has been trained extensively on large data sets and performed well in identifying human body features, this research adopted it for real-time 2D human feature detection. The pretrained RMPE yields 17 default key joint points representing human body...
parts (Figure 5). The key joint points and their corresponding human body parts are listed in Table 2.

The 2D body skeletons extracted from the images using RMPE are represented in Equations 2, 3, and 4:

\[ E = \{ S_i, S_2, ..., S_k \} \]
\[ S = \{ j, 1 \leq i \leq m \}, 0 \leq m \leq 16, S \in E \]
\[ j = (x_i, y_i), 0 \leq i \leq m, \]

where \( E \) is a set of human body skeletons \( S_j \) \( j \in \{1, 2, ..., k\} \) of \( k \) people detected by RMPE in the image. Each skeleton \( S \) is a set of 2D joint points \( j \) \( j \in \{1, 2, ..., m\} \) that contain 2D coordinates \((x_i, y_i)\), which correspond to the left-view image. \( m \) is the total number of body parts listed in Table 2. Each pixel in a 3D map contains both 2D image coordinates and 3D coordinates.

The 2D body skeletons were converted to 3D body features by finding the 3D coordinates corresponding to the 2D joint points from the 3D map using the 2D image coordinates as the index. Thus, a set of 3D body features containing depth information was derived at this stage and saved in the queue for further analysis. These 3D body features were used to evaluate human kinematics, which will be discussed in the next section.

### Analysis and Visualisation of 3D Human Kinematics

Threads 1 and 2 worked continuously with the frames loaded from the stereo camera. The 3D body features extracted from a series of stereo camera frames then facilitated the kinematic analysis of a 3D human body over time in thread 3. This study focused on typical 3D human kinematics, including the velocity of movement (moving speed and direction), step length, knee flexion angle, and arm swing angles. Table 3 lists detailed descriptions of the considered human kinematics based on the default 17 key joint points (Table 2).

The human center of mass is maintained or altered close to the midpoint of the left and right hips (Vlutters et al. 2016). Therefore, the midpoint of the left and right hips was used to calculate moving speed and direction. The moving speed was calculated based on the 3D coordinates of the midpoint at the initial and final positions of a person’s movement during a time interval. The moving direction was treated based on trigonometry that movement can be in any direction in a 360° arc starting from the direction in which the person faces the camera. The 360° were divided into groups to represent different directions. The moving direction in this system was classified into four directions: forward, backward, left, and right (Figure 6). In Figure 6, \( Pi \) \( i = 0 \) indicates the possible initial position, and \( Pi \) \( i = 1, 2, 3, 4 \) illustrates the possible final positions in each direction in the next frame. The direction was determined by calculating the angle \( \theta_i \) between the vector from an initial position to a final position and the XY-plane of the camera system based on the 3D coordinates of two hips. The step length was expressed as the vector length from one ankle to the other. The system calculated the step length using the 3D coordinates of both ankles while calculating the direction and speed of movement.
displayed all of the results by thread 4 in a window. As shown in Figure 8, the background is the stitching image with the left-view image of the camera and colored disparity map. Red colors on the disparity map indicate objects closer to the camera, whereas darker blue colors represent objects further away from the camera. Each joint is connected by different-colored lines. The distance of each body joint was loaded from 3D information in the queue and drawn on the left side of the background beside each body joint using 2D coordinates. All kinematic results were loaded from the queue and displayed on the colored disparity map for real-time monitoring of human locomotion.

**System Implementation and Evaluation**

**Hardware Configuration of the System**

The camera system used in this research is a ZED camera, which includes a stereo pair of RGB cameras of the same model on the same mainboard. The baseline between the two cameras is 12 cm, and each camera has a horizontal FOV of 90° and a vertical FOV of 60°. The left and right cameras each have a focal length of 5.6 mm. The image resolution is 672 × 376 pixels for each camera, with a pixel size of 4 µm. The camera system was calibrated by the manufacturer. The camera interior orientation parameters, including the focal length, offset of the principal point, and lens distortions, and a fundamental matrix defining the relative orientation of the stereo cameras, are provided and ready for use. We used a local coordinate system in the experiment with the origin at the perspective center of the left camera, X-axis along the baseline, Y-axis pointing downwards, and Z-axis pointing to the range direction (see Figure 3). The camera system was run on a computer equipped with two NVIDIA RTX 2080Ti graphics cards, 64 GB of RAM, and two 12-core CPUs.

**Evaluation of the System Capacity**

The capabilities of the developed system were evaluated by assessing the processing rate and effective detection distance of a person moving in front of the stereo camera. During the assessment, 6000 frames were captured within 300 seconds.
The implementation of all threads reached ~18 fps or above with an image resolution of 672 × 376 pixels. The average processing rate of this system was 17.8 fps. Figure 9a illustrates the processing time of each frame from thread 1 to thread 4. According to Figure 9a, the processing rate sometimes exceeded 20 fps. This occurred when the person moved so fast that a ghosting effect appeared on the corresponding frames, or the illumination was so weak that the person barely disappeared from the screen. As a result, RMPE failed to extract the 2D human skeletons in the above situations. In response to such situations, thread 2 skipped the current frame and processed the next frame directly, resulting in a moderate uplift in frame rate. The entire assessment took 6000 frames (Figure 9a), and in general, the developed system achieved real-time processing during the assessment.

The system achieved an effective measurement distance of ~15 m, based on assessing a person moving back and forth from near to far along the optical axis of the left camera. During the evaluation, when the person left the camera view and returned along the same path, the system recorded all distance values from the person’s waist, defined as the midpoint between the left and right hips, during this movement. As shown in Figure 9b, when the person moved ~1.5 m, the system was able to extract the 3D body features of the left and right hips and started to compute the corresponding 3D coordinates. When the person moved more than 15.7 m away from the camera, the system could not measure the distance because the person became too small on the screen to be detected by the RMPE. As the person began to move back towards the camera and moved within 14.2 m, the system was able to extract the 3D human body features again and simultaneously calculated the distance until the person moved to a distance of less than 1.1 m from the camera. Measurements were unstable starting at 14.2 m, whereas the dead zone for close-range measurements was from ~1.1 m to ~1.5 m. Thus, the effective measurement range was ~1.5 m to ~15 m, which covers a large scale of scenes.

### Accuracy Evaluation of the 3D Human Body Kinematics

#### Distance Accuracy

To evaluate the accuracy of the distance measurements achieved by the system over a specific resolution, we had a person standing still in front of the camera at different distances (Figure 10). The distance accuracy was assessed by comparing the measured distances between the person and the camera to the ground truth. As shown in Figure 10b, the system captured 1000 frames of the person when the person was standing still in front of the camera at distances of 2.3 m,
4.1 m, 12 m, and 15 m. Table 4 lists the measurement averages. According to Table 4, the measurements of the system were close to the ground truth. When the person was 2.3 m and 4.1 m away from the camera, the respective root-mean-square errors (RMSEs) were 0.4 cm and 2.6 cm, respectively. The errors were 0.2% and 0.6%, respectively. As the person moved to 12 m, the measurements began to become unstable. The RMSE increased to 8.7 cm, and the error became 0.7%. When the person stood 15 m away from the camera, the measurements were even more erratic. The RMSE increased to 47.9 cm, and the error rose to 3.2%. Because the effective measurement range was ~1.5 m to 15 m (see the section “Evaluation of the System Capacity”), the measurements at 15 m were not detected on each frame. Overall, this system provided 3D human body measurements with a geometric accuracy of better than 1% of the distance within the distance range of 12 m.

**Accuracy of Human Body Kinematics**

The human moving direction was assessed by recording a person moving in four directions relative to the camera: left, right, forward, and backward. Figure 11 shows the method to evaluate the moving direction. Figure 11a shows the initial

### Table 4. Assessment of system measurement accuracy.

<table>
<thead>
<tr>
<th>Distance (m)</th>
<th>Mean of Measurements (m)</th>
<th>RMSE (m)</th>
<th>Error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.3</td>
<td>2.3</td>
<td>0.004</td>
<td>0.2</td>
</tr>
<tr>
<td>4.1</td>
<td>4.1</td>
<td>0.026</td>
<td>0.6</td>
</tr>
<tr>
<td>12.0</td>
<td>12.1</td>
<td>0.087</td>
<td>0.7</td>
</tr>
<tr>
<td>15.0</td>
<td>15.1</td>
<td>0.479</td>
<td>3.2</td>
</tr>
</tbody>
</table>

### Table 5. Statistic results of moving direction identification.

<table>
<thead>
<tr>
<th>Figure 11 panel</th>
<th>Expected Behavior</th>
<th>Test Times</th>
<th>Average $\theta$ (°)</th>
<th>Average Speed (cm/s)</th>
<th>Accuracy of Correct Identification (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>Standing still</td>
<td>30</td>
<td>0.2</td>
<td>0</td>
<td>93</td>
</tr>
<tr>
<td>(b)</td>
<td>Moving left</td>
<td>30</td>
<td>88.1</td>
<td>52</td>
<td>87</td>
</tr>
<tr>
<td>(c)</td>
<td>Moving right</td>
<td>30</td>
<td>-89.7</td>
<td>55</td>
<td>90</td>
</tr>
<tr>
<td>(d)</td>
<td>Moving forward</td>
<td>30</td>
<td>-10.2</td>
<td>43</td>
<td>83</td>
</tr>
<tr>
<td>(e)</td>
<td>Moving backward</td>
<td>30</td>
<td>-2.4</td>
<td>61</td>
<td>93</td>
</tr>
</tbody>
</table>

Figure 11. Results of monitoring human movement direction. The direction of movement is determined relative to the position of the camera. (a) The initial position of the human. (b) Identified result of moving left. (c) Identified result of moving right. (d) Identified result of moving forward. (e) Identified result of moving backward.
position of the person, and Figures 11b–11e display the monitoring results of the person moving in four different directions, with moving speed computed in real time.

The identification of moving direction was performed following the geometry shown in Figure 6. We evaluated each direction and repeated the measurements 30 times for each direction. The results are listed in Table 5. According to the results, the identified moving direction is generally consistent with the expected behavior in each direction, with an accuracy of over 83%. Figure 11 shows examples of the results.

Table 6 and Figure 12 present the kinematic analysis results, including step length, knee angles, elbow angles, and upper-arm swing angles, measured and recorded from 1000 frames by letting a person stand in front of the camera a while. Ground truth data were manually measured by a ruler for the step length and a protractor for the angles. An RMSE of 0.3 cm

<table>
<thead>
<tr>
<th>Kinematic Application</th>
<th>Mean of Measurements</th>
<th>Ground Truth</th>
<th>RMSE (cm)</th>
<th>Error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step length (cm)</td>
<td>32.6</td>
<td>33.1</td>
<td>0.3</td>
<td>0.8</td>
</tr>
<tr>
<td>Knee angle (°)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td>169.7</td>
<td>176.0</td>
<td>6.3</td>
<td>3.6</td>
</tr>
<tr>
<td>Right</td>
<td>170.4</td>
<td>176.0</td>
<td>5.7</td>
<td>3.2</td>
</tr>
<tr>
<td>Elbow angle (°)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td>164.1</td>
<td>161.0</td>
<td>5.4</td>
<td>3.4</td>
</tr>
<tr>
<td>Right</td>
<td>166.4</td>
<td>160.0</td>
<td>7.1</td>
<td>4.4</td>
</tr>
<tr>
<td>Upper-arm angle (°)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td>33.6</td>
<td>35.0</td>
<td>2.6</td>
<td>7.3</td>
</tr>
<tr>
<td>Right</td>
<td>32.9</td>
<td>31.0</td>
<td>2.3</td>
<td>7.5</td>
</tr>
</tbody>
</table>

Figure 12. Kinematic analysis of system measurements. (a) 1000-frame measurements of step length, knee flexion angles, and arm swing angles. (b) The system-measured kinematic results of a person standing still in front of the camera for a moment.
is calculated for measured step length with an error of 0.8%. The left and right knee angles fluctuate slightly around 170°, with RMSEs of 6.3° and 5.7°, respectively. The error remained at about 3%. The elbow and upper-arm angle measurements are unstable due to the high illumination at their positions on the image. This uncertainty results in an inaccuracy of the RMPE in extracting the 2D features and 3D feature conversion. The mean of the elbow angle measurements on both sides hovers at 164.1° and 166.4°, respectively. The RMSE was 5.4° and 7.1°, respectively, with an error of less than 5% for both. These values are relatively stable overall. The RMSE and errors of the left and right upper-arm angles are 2.6° (error 7.3%) and 2.3° (error 7.5%). This result indicates that the measured angles were slightly discrepant relative to the ground truth value.

Conclusions and Discussion

This paper proposed a novel real-time photogrammetric system for 3D human body feature extraction with potential applications for human kinematics. The run-time frame rate of all frameworks, including 3D map generation, 2D and 3D human body feature extraction, and human kinematic analysis, was improved by multi-threading on the GPU and CUDA programming on the GPU. The 3D map was derived from disparities using the GPU-accelerated SGM method and photogrammetry method. Human body features in 2D and 3D were extracted using the deep-learning-based RMPE method and were run in an individual thread. Several geometric models were introduced as an example of human kinematic analysis. The experimental results presented in this paper quantitatively evaluate the efficiency and accuracy of each measurement for human kinematic analysis. The process rate (pose framerate) reached ~18 fps. The effective detection distance reached 15 m, with a geometric accuracy of better than 1% of the distance within a range of 12 m. The accuracy for real-time measurement of human body kinematics ranged from 0.8% to 7.5%. Our system achieved large-scale 3D human body feature detection. The integration of deep learning methods let the system accurately recognize 3D human body features for human kinematic analysis. With the help of multi-threading and GPU-acceleration technology, this system improved the running framerate and achieved real-time 3D human monitoring at a large scale.

There are some limitations in the experiments. The RMPE failed to detect the 2D human features when the person was moving very fast (a ghosting effect appeared on the screen) or when the illumination was dark (the person almost disappeared from the screen). Similarly, the light intensity in the environment was not constant, and the SGM did not accurately obtain the disparity value in a very high-lighting environment, such as an area near a lamp, or low-lighting environments, such as shadows. The 3D information was not extracted in these cases. Moreover, 3D body features were not extracted if a person was standing more than 15 m from the camera. At this distance, the person was so small in the image that the 2D human detection algorithm was unable to extract human skeletons. These problems can be improved by optimizing the algorithms to support higher image resolutions. The clearer outline of a person in a higher-resolution image allows the deep learning method to recognize the body features at farther distances. It should be noted that the current system is only able to process image sequences of a resolution of 672 × 376 pixels in real-time. With proper optimization of the software in our future works, real-time processing of image sequences of higher resolutions can be expected. It should also be noted that the moving directions that can be identified in the current system only allow four main directions. The algorithms will be further improved in our future work to allow the identification of more sophisticated moving directions.

In this study, although several applications of 3D human kinematics, including joint angles, movement directions, etc., were selected for demonstration and evaluation of their accuracy, the system was not limited to these applications. We expect that this study provides an insight into the potential applications of real-time 3D photogrammetry. We also hope that this system would be integrated into a portable device that could extend real-time photogrammetry to a wider range of scientific fields and industries in the future.

Acknowledgments

This work was supported by grants from the Hong Kong Polytechnic University (Project No. 1-ZVN6) and the National Natural Science Foundation of China (Project No. 41671426).

References


Benefits of an ASPRS Membership

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

*quantity depends on membership level
Modeling Hyperhemispherical Points and Calibrating a Dual-Fish-Eye System for Close-Range Applications

Letícia Ferrari Castanheiro, Antonio Maria Garcia Tommaselli, Adilson Berbevleri, Mariana Batista Campos, and José Marcato Junior

Abstract
Omnidirectional systems composed of two hyperhemispherical lenses (dual-fish-eye systems) are gaining popularity, but only a few works have studied suitable models for hyperhemispherical lenses and dual-fish-eye calibration. In addition, the effects of using points in the hyperhemispherical field of view in photogrammetric procedures have not been addressed. This article presents a comparative analysis of the fish-eye models (equidistant, equisolid-angle, stereographic, and orthogonal) for hyperhemispherical-lens and dual-fish-eye calibration techniques. The effects of adding points beyond 180° field of view in dual-fish-eye calibration using stability constraints of relative orientation parameters are also assessed. The experiments were performed with the Ricoh Theta dual-fish-eye system, which is composed of fish-eye lenses with a field of view of approximately 190° each. The equisolid-angle model presented the best results in the simultaneous calibration experiments. An accuracy of approximately one pixel in the object space units was achieved, showing the potential of the proposed approach for close-range applications.

Introduction
Spherical panoramas—that is, a 360° field of view (FOV)—have been widely used for many mobile mapping applications, such as robotic navigation and driverless vehicles (Campos et al. 2018a; Perfetti et al. 2018; Meegoda et al. 2019; Sun and Zhang 2019; Gao et al. 2020). These applications have motivated research focusing on the design of high-performance imaging systems covering a full-spherical FOV (Pernechele 2016; Song et al. 2018). However, to achieve high geometric accuracy, calibration techniques suitable to these systems are required (Aghayari et al. 2017; Campos et al. 2018b; Jarron et al. 2019; Khoramshahi et al. 2019). Attractive options are cameras with dual-fish-eye lenses arranged in a compact structure in a back-to-back position, enabling full-spherical coverage (e.g., Ricoh Theta S, GoPro Fusion, Xiaomi Mi Sphere 360). Due to this compact arrangement, the cameras’ perspective centers are located very close to each other. These cameras have an overlapping FOV between images, which is essential in creating panoramic images without blind zones (Pernechele 2016; Song et al. 2018). This common FOV between cameras is also relevant in creating common tie points in the camera calibration step, which can improve photogrammetric procedures such as camera orientation. These points present small parallactic angles (<1°), where the rays are approximately aligned along the trajectory, in mobile mapping applications. Civera et al. (2008) and Schneider and Förstner (2013) have noted that observations with such a small parallax improve the estimation of attitude angles.

Despite these advantages, a proper mathematical model differing from the well-known perspective model is required to model a wider-FOV camera. Recent works have focused on mathematical models for camera calibration (Aghayari et al. 2017; Campos et al. 2018b; Gao et al. 2020) and close-range applications (Campos et al. 2018a; Perfetti et al. 2018; Campos et al. 2019) using omnidirectional systems with hyperhemispherical lenses. However, points beyond 180° FOV are usually excluded from photogrammetric procedures such as camera calibration (Campos et al. 2018b) and 3D reconstruction (Perfetti et al. 2018; Campos et al. 2019), since some mathematical models do not cope with these points. Campos et al. 2018b presented a methodology to calibrate a dual-fish-eye system based on high accuracy and efficiency, and some of those systems have drawbacks restricting their use. Catadioptric systems have been used in CRP mobile mapping systems (Marcato et al. 2016; Filin et al. 2020); however, the images usually have a blind area caused by camera occlusion, and the system calibration process can be quite complex. The use of a rotating camera to produce a panoramic image is not suitable for mobile applications, due to the capture-time delay between two shots (Won et al. 2019). A multi-camera system covering a 360° FOV can offer many advantages compared to other omnidirectional systems.

Multi-camera systems can be designed with perspectives or fish-eye lenses. Perspective cameras have a narrower FOV, and several cameras are needed to cover a full-spherical panorama (e.g., Ladybug5 and Panono Pro 360°), which can result in higher costs and complex camera synchronization compared to fish-eye lens systems (Tommaselli et al. 2013; Lichti et al. 2013; Jarron et al. 2019; Khoramshahi et al. 2019). Attractive options are cameras with dual-fish-eye lenses arranged in a compact structure in a back-to-back position, enabling full-spherical coverage (e.g., Ricoh Theta S, GoPro Fusion, Xiaomi Mi Sphere 360). Due to this compact arrangement, the cameras’ perspective centers are located very close to each other. These cameras have an overlapping FOV between images, which is essential in creating panoramic images without blind zones (Pernechele 2016; Song et al. 2018). This common FOV between cameras is also relevant in creating common tie points in the camera calibration step, which can improve photogrammetric procedures such as camera orientation. These points present small parallactic angles (<1°), where the rays are approximately aligned along the trajectory, in mobile mapping applications. Civera et al. (2008) and Schneider and Förstner (2013) have noted that observations with such a small parallax improve the estimation of attitude angles.

Despite these advantages, a proper mathematical model differing from the well-known perspective model is required to model a wider-FOV camera. Recent works have focused on mathematical models for camera calibration (Aghayari et al. 2017; Campos et al. 2018b; Gao et al. 2020) and close-range applications (Campos et al. 2018a; Perfetti et al. 2018; Campos et al. 2019) using omnidirectional systems with hyperhemispherical lenses. However, points beyond 180° FOV are usually excluded from photogrammetric procedures such as camera calibration (Campos et al. 2018b) and 3D reconstruction (Perfetti et al. 2018; Campos et al. 2019), since some mathematical models do not cope with these points. Campos et al. 2018b presented a methodology to calibrate a dual-fish-eye system based on high accuracy and efficiency, and some of those systems have drawbacks restricting their use. Catadioptric systems have been used in CRP mobile mapping systems (Marcato et al. 2016; Filin et al. 2020); however, the images usually have a blind area caused by camera occlusion, and the system calibration process can be quite complex. The use of a rotating camera to produce a panoramic image is not suitable for mobile applications, due to the capture-time delay between two shots (Won et al. 2019). A multi-camera system covering a 360° FOV can offer many advantages compared to other omnidirectional systems.

Multi-camera systems can be designed with perspectives or fish-eye lenses. Perspective cameras have a narrower FOV, and several cameras are needed to cover a full-spherical panorama (e.g., Ladybug5 and Panono Pro 360°), which can result in higher costs and complex camera synchronization compared to fish-eye lens systems (Tommaselli et al. 2013; Lichti et al. 2013; Jarron et al. 2019; Khoramshahi et al. 2019). Attractive options are cameras with dual-fish-eye lenses arranged in a compact structure in a back-to-back position, enabling full-spherical coverage (e.g., Ricoh Theta S, GoPro Fusion, Xiaomi Mi Sphere 360). Due to this compact arrangement, the cameras’ perspective centers are located very close to each other. These cameras have an overlapping FOV between images, which is essential in creating panoramic images without blind zones (Pernechele 2016; Song et al. 2018). This common FOV between cameras is also relevant in creating common tie points in the camera calibration step, which can improve photogrammetric procedures such as camera orientation. These points present small parallactic angles (<1°), where the rays are approximately aligned along the trajectory, in mobile mapping applications. Civera et al. (2008) and Schneider and Förstner (2013) have noted that observations with such a small parallax improve the estimation of attitude angles.

Despite these advantages, a proper mathematical model differing from the well-known perspective model is required to model a wider-FOV camera. Recent works have focused on mathematical models for camera calibration (Aghayari et al. 2017; Campos et al. 2018b; Gao et al. 2020) and close-range applications (Campos et al. 2018a; Perfetti et al. 2018; Campos et al. 2019) using omnidirectional systems with hyperhemispherical lenses. However, points beyond 180° FOV are usually excluded from photogrammetric procedures such as camera calibration (Campos et al. 2018b) and 3D reconstruction (Perfetti et al. 2018; Campos et al. 2019), since some mathematical models do not cope with these points. Campos et al. 2018b presented a methodology to calibrate a dual-fish-eye system based

Letícia Ferrari Castanheiro, Antonio Maria Garcia Tommaselli, and Adilson Berbevleri are with the Department of Cartography, São Paulo State University, School of Technology and Sciences, Presidente Prudente, Brazil (castanheirof@gmail.com).

Mariana Batista Campos is with the Department of Remote Sensing and Photogrammetry, Finnish Geospatial Research Institute, Masala, Finland.

José Marcato Junior is with the Faculty of Engineering, Architecture and Urbanism and Geography, Federal University of Mato Grosso do Sul, Campo Grande, Brazil.

Contributed by Bo Wu, August 24, 2020 (sent for review December 10, 2020; reviewed by Sagi Filin and Mohammad Omidiziarandi).
on the equidistant model. They removed the hyperhemispherical points from the camera calibration process to avoid an incorrect mapping with the equidistant model, which occurs due to the arctan function in this model. Perfetti et al. (2018) discussed the use of fish-eye optics in camera calibration and 3D reconstruction. In the camera calibration, the peripheral points were removed to model the fish-eye distortions using the conventional perspective camera model and its classic radial distortion coefficients. For 3D reconstruction, marginal areas of the images were discarded due to the decreased resolution, which generates a less noisy dense point cloud.

Few studies have proposed mathematical models that can cope with a hyperhemispherical FOV. Song et al. (2018) designed a dual-fish-eye system that was calibrated using a generic polynomial model with lens distortions directly included in the set of parameters. Gao et al. (2020) also used a polynomial model to calibrate their developed dual-fish-eye system composed of 245° FOV lenses. Similarly, Scaramuzza et al. (2006) proposed a generic model based on Taylor-series expansion, which was tested in hyperhemispherical lenses by Won et al. (2019). However, the degree of polynomial parameters increases with the camera’s FOV, because more coefficients are needed to model the distortions (Hughes et al. 2010; Usenko et al. 2018). Generic models based on spherical projection were proposed by Khomutenko et al. (2016) and Usenko et al. (2018) to avoid high-order polynomials. For the same purpose, Pernechele (2016) designed a hyperhemispherical lens following a mapping function with a few parameters. However, generic models can lead to overfitting parameter estimation (Abraham and Förstner 2005, p. 279). Rigorous fish-eye models (equidistant, equisolid-angle, stereographic, or orthogonal) are based on physical imaging principles (Schneider et al. 2009; Hughes et al. 2010), which are a better choice to avoid polynomial models and instabilities in the calibration procedure. Modeling hyperhemispherical lenses with rigorous fish-eye models has not been fully explored, and to the best of our knowledge, no previous work has focused on this feature.

The novel contribution of this work is to present a comparative analysis of fish-eye models for the suitable modeling of hyperhemispherical points that enables the inclusion of these points in a dual-fish-eye calibration. Rigorous fish-eye models were compared and assessed in single- and dual-camera calibration using the Ricoh Theta dual-fish-eye system. The effect of including hyperhemispherical points was assessed in a simultaneous camera calibration based on stability constraints of relative orientation parameters (ROPs). Statistical analyses of the estimated parameters and observation residuals in the camera calibration were performed for each fish-eye model.

<table>
<thead>
<tr>
<th>Models</th>
<th>Geometry of projection</th>
<th>Equations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equidistant</td>
<td>$r_d = ac$</td>
<td></td>
</tr>
<tr>
<td>Equisolid-angle</td>
<td>$r_d = 2c \sin(\alpha/2)$</td>
<td></td>
</tr>
<tr>
<td>Stereographic</td>
<td>$r_d = 2c \tan(\alpha/2)$</td>
<td></td>
</tr>
<tr>
<td>Orthogonal</td>
<td>$r_d = c \sin(\alpha)$</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1. Mathematical models for fish-eye-lens cameras (adapted from Abraham and Förstner 2005; Hughes et al. 2010, Marcato et al. 2015; Campos et al. 2018b).
lenses (Schneider et al. 2009; Tommaselli et al. 2014; Marcato et al. 2015), mainly for points out of the image center. For this reason, only fish-eye models were compared in this article.

Figure 1 presents the mathematical models and image-projection geometry for rigorous fish-eye models (equidistant, equisolid-angle, stereographic, and orthogonal), in which $r_j$ is the radial distance in the image, $\alpha$ is the incident angle, $\beta$ is the refracted angle, $(x_0, y_0)$ are the principal-point coordinates, $Ax$ and $Ay$ are distortion models, and $c$ is the principal distance. The relationship of the point coordinates in the photogrammetric system $(X_0, Y_0, Z_0)$ with the corresponding ground coordinates $(X, Y, Z)$ requires a similarity transformation as a function of the perspective-center (PC) coordinates $(X_C, Y_C, Z_C)$ and Euler angles $(\omega, \varphi, \kappa)$ that are embedded in a rotation matrix (Schenk 2005). A detailed geometric description of these fish-eye models is given by Hughes et al. (2010) and Marcato et al. (2015).

Limitations for some rigorous fish-eye projections in terms of modeling hyperhemispherical lenses have been reported in the literature (Hughes et al. 2010; Pernechele, 2016, Campos et al. 2018b), leading to the removal of hyperhemispherical points from the photogrammetric process to obtain consistent results. These limitations are illustrated in Figure 2. Figure 2a illustrates a camera (PC) and four points in the object space, considering a right-handed reference system $(X, Y, Z)$. Points 1, 2, and 3 are located in the hemispherical FOV, and point 4 is in the hyperhemispherical FOV. A principal distance of 1.43 mm (the same as for the Ricoh Theta system) was assumed for the camera. The points, which have the same $Y$-coordinate, were projected to the image plane $(x, y)$ following the equidistant, orthogonal, equisolid-angle, and stereographic models.

Figure 2b shows the mapping problems for observations beyond a 180° FOV using the equidistant model in a dual-fish-eye camera calibration process, as mentioned by Campos et al. 2018b. Points 3 and 4 should be projected to the right side of the image, next to point 1. In Figure 2c (orthogonal model), an incorrect mapping for the hyperhemispheric point 4 can be seen; it should be projected to the right side of point 3. Therefore, orthogonal projection can be considered a suitable model only for hemispherical fish-eye lenses, due to the perpendicular projection of the points (Hughes et al. 2010; Pernechele, 2016). The equisolid-angle (Figure 2d) and stereographic (Figure 2e) models reprojected the four points correctly. Therefore, further investigations of hyperhemispherical lens modeling are still required.

**Fish-Eye Model Assessment**

**Data Set**

The Ricoh Theta S dual-fish-eye system is composed of two fish-eye lenses, each with a 190° FOV (hyperhemispherical geometry) and a nominal principal distance of 1.31 mm (Theta Developers 2017). Dual-fish-eye frames, with a pixel size of 0.005 mm, were taken with the Ricoh Theta S dual-fish-eye camera in video mode in a 360° calibration field composed of 160 ArUco targets (Campos et al. 2018b). The preprocessing from video to individual fish-eye frames is detailed by Campos et al. 2018b.

The ArUco targets were automatically detected (Garrido-Jurado et al. 2014) with subpixel accuracy (Tommaselli et al. 2014); therefore, a standard deviation of 0.5 pixels for each component (column and row) was assumed. Ground coordinates of the four corners of each ArUco target were determined by a bundle adjustment with a large number of images from a high-resolution camera (Sony NEX), with further observations coming from total station observations. An average estimated accuracy of 5 mm was obtained in a previous bundle adjustment (Campos et al. 2018b, p. 251), which is compatible with the pixel size in the object space units of the fish-eye images used. Therefore, all ground coordinates of ArUco targets were considered as weighted constraints in the bundle adjustment, with a standard deviation of 0.005 m. Standard deviations of the initial exterior orientation parameters (EOPs) were set to 0.5 m and 10° for camera position and attitude, respectively.

Twenty video frames from sensor 1 covering the 360° camera calibration field were used to assess the fish-eye models in the single-camera calibration (Experiment A). Then a dual-fish-eye camera calibration (Experiment B) was performed with 13 images from each sensor (total of 26 images). Convergent and rotated images, with varying $\omega$, $\varphi$, and $\kappa$ angles (the latter with 0°, 90°, 180°, and –90° values), were taken in different positions to minimize the correlations between the calibration parameters (Fraser 1997). The maximum distance between stations was 4 m. Figure 3a shows the Ricoh Theta S dual-fish-eye camera, and Figure 3b shows examples of the Ricoh Theta video frames taken in the 360° calibration field, highlighting the overlapping field between sensor 1 (S1) and sensor 2 (S2), which is beyond the circle. Figure 3c presents a top view of the network geometry of the images used in the dual-fish-eye camera calibration, with ground targets appearing. Targets on the walls were not represented in the top view in Figure 3c.
To assess the results of camera calibration using points in the hyperhemispherical image field, two sets of points were considered in the experiments. The first data set considered only points located in the hemispherical image field (α < 90°)—that is, points inside the circle presented in Figure 3b. This data set was named hemispheric (hem). The second data set, named hyperhemispheric (hyp), included the hemispheric data set and added points beyond the circle. Both data sets were tested in two experiments: (A) hyperhemispherical single-lens camera calibration and (B) dual-fish-eye camera calibration.

Experiment A: Hyperhemispherical-Lens Camera Calibration

In Experiment A, a single Ricoh Theta S camera (sensor 1) was calibrated using rigorous fish-eye models (equidistant, stereographic, equisolid-angle, and orthogonal) combined with the Conrady–Brown distortion model. The models were implemented in C/C++ on calibration-with-multiple-cameras software (Tommaselli et al. 2010; Marcato et al. 2015). Calibration parameters were estimated based on the unified approach to least squares with constraints (Mikhail and Ackerman 1976). Equations 1 and 2 present the distortion model used, in which \( K_x, K_y, \) and \( K_2 \) are symmetric radial distortion coefficients (Brown 1971), \( P_x \) and \( P_y \) are decentering distortion coefficients (Conrady 1919; Brown 1971), and \( A \) and \( B \) are affinity parameters (Fraser 1997):

\[
\Delta x = x(K_x r^2 + K_y r^4 + K_r r^6) + P_x r^2 + 2 P_y x y + A x + B y \quad (1)
\]

\[
\Delta y = y(K_x r^2 + K_y r^4 + K_r r^6) + P_x r^2 + 2 P_y x y . \quad (2)
\]

The significance of the estimated distortion parameters was first analyzed. Then the camera calibration was repeated, considering only distortion parameters in which the effects on the image limits were higher than the image measurement error (0.5 pixel). The bundle-adjustment trials were performed with 252 measured points in a hemispherical FOV (hem data set: Experiment A.1) and 308 measured points in an all-image FOV (hyp data set: Experiment A.2). The outliers were manually removed based on residual analysis of the observations.

The estimated standard deviation of unit weight \( \hat{e}_n \) was compared with the a priori value \( e_n \), which was set to 1 (Vanček 1973, p. 202). Statistical analysis of the observation residuals (mean, standard deviation, and root-mean-square error [RMSE]) in the resultant component are also presented. Finally, five control distances among ArUco targets were directly measured with a caliper (precision of 0.2 mm, measuring lengths up to 2 m) in the 360° calibration field and compared with the calculated distances that were obtained from the ground coordinates estimated in the bundle adjustment.

Experiment B: Dual-Fish-Eye Camera Calibration

The models that achieved the best results in Experiment A.2 were used in the calibration-with-multiple-cameras software to calibrate both Ricoh Theta sensors in a simultaneous bundle adjustment. Campos et al. 2018b verified improvements from using constraints on the stability of ROPs (Tommaselli et al. 2013; Lichti et al. 2015) in dual-fish-eye camera calibration. ROPs can be considered stable during image acquisition but admitting small random variations. Therefore, the constraint equations can be introduced in the bundle adjustment, considering the null differences between the ROPs as pseudo-observations, with a variance obtained by covariance propagation from admitting variations in the ROPs. Detailed information about stability constraints on self-calibrating bundle adjustment are given by Tommaselli et al. (2013).

In this work, stability constraints on the base elements and relative rotation angles were considered in the bundle adjustment of Experiment B. A standard deviation of 0.0005° for each relative rotation angle and a variation of 0.001 mm between the camera’s base elements during the image acquisition were set. After the camera calibration, the base elements and relative rotation angles were determined for each pair of ROPs. The base elements relative to camera 1 (C1) were determined as presented in Equation 3. The relative orientation matrix \( R_{h2} \) was calculated from the rotation matrices of both cameras \( R_{h1} \) and \( R_{h2} \), as shown in Equation 4 (Tommaselli et al. 2013; Campos et al. 2018b):

\[
B_x = R_{h1} \begin{bmatrix} x^2 \ y^2 \ z^2 \ \end{bmatrix}
\]

\[
B_y = R_{h1} \begin{bmatrix} y^2 \ z^2 \ x^2 \ \end{bmatrix}
\]

\[
B_z = R_{h1} \begin{bmatrix} z^2 \ x^2 \ y^2 \ \end{bmatrix}
\]

\[
R_{h2} = R_{h1}^{-1} \begin{bmatrix} z^2 \ x^2 \ y^2 \ \end{bmatrix} . \quad (3)
\]

For this experiment, the hemispheric data set was composed of 320 observations (Experiment B.1) from images of both sensors, and the hyperhemispheric data set consisted of 368 measured points (Experiment B.2). Statistical analyses of the observation residuals were performed as in Experiment A, including a comparison of the interior orientation parameters (IOPs) and ROPs estimated with the hemispheric and hyperhemispheric data set, which was extended for the estimated EOP standard deviations. The accuracy of the bundle adjustment was assessed by discrepancies in the eight directly measured distances well distributed in the 360° calibration field.

An important issue in camera calibration is the correlation coefficient between IOPs and ROPs, which was also evaluated by considering five images. This assessment was performed only for the fish-eye model that achieved the best results in the simultaneous camera calibration. The comparison was performed for the correlation coefficients estimated in Experiments B.1 and B.2. A further analysis was conducted by comparing the results of calibration with and without relative orientation constraints.

Results and Discussion

Experiment A: Hyperhemispherical-Lens Camera Calibration

The bundle-adjustment solution based on the equidistant model failed with the hyperhemispheric data set. The main reason for this lack of a solution is the points in the hyperhemispherical FOV that have the effect of gross errors, as shown in Figure 2. Therefore, the significance of the distortion parameters was analyzed only for the other models. Table 1 presents the effects (in pixels) in the image coordinates (x, y) in the corner of the sensor achieved for the equisolid-angle, stereographic, and orthogonal projection models.

<table>
<thead>
<tr>
<th>Model</th>
<th>Radial Symmetric x</th>
<th>Decentering x</th>
<th>Affinity x</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equisolid-angle</td>
<td>–545.09</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Stereographic</td>
<td>–2275.39</td>
<td>1.04</td>
<td>0.06</td>
</tr>
<tr>
<td>Orthogonal</td>
<td>622.24</td>
<td>0.00</td>
<td>0.01</td>
</tr>
</tbody>
</table>

A significant radial symmetric distortion is one of the characteristics of fish-eye images due to the large FOV, which explains the high values presented in Table 1. Decentering distortion effects were much smaller, but still higher than 0.5 pixel (measurement error), for the equisolid-angle and stereographic models. The effects of affinity distortion are lower than 0.5 pixel for all projection models, which was also observed by Campos et al. 2018b in a calibration based on the equidistant model with only observations in the hemispheric...
stereographic projection was considered suitable and were used to perform the calibration of the Ricoh Theta system in a simultaneous bundle adjustment.

**Experiment B: Simultaneous Dual-Fish-Eye Camera Calibration**

Since the equisolid-angle and stereographic fish-eye projections presented suitable results to calibrate a Ricoh Theta hyperhemispherical sensor (Experiment A.2), a simultaneous calibration by bundle adjustment was performed based on both models. Thirteen images taken by each Ricoh Theta sensor in the 360° calibration field were used. Figure 4 shows the mean, standard deviation, and RMSE of image-coordinate residuals in the resultant component, as well as the estimated standard deviations of unit weight ($\hat{\sigma}_c$) for the equisolid-angle and stereographic models with both data sets when simultaneously processing images from both cameras with bundle adjustment.

The estimated standard deviations of unit weight for both the equisolid-angle and stereographic projections were smaller than the a priori value ($\sigma_0 = 1$). The addition of the hyperhemispheric observations increased the image residuals in the simultaneous bundle adjustment based on the stereographic projection (Figure 4), and consequently $\hat{\sigma}_c$ was also higher. This increase in the residual observations can be explained by the lower resolution in the border of the fish-eye images. Therefore, a different stochastic treatment of the observations can be required due to the nonuniform resolution of fish-eye images, which will be performed in future works. On the other hand, the estimated standard deviation of unit weight for the equisolid-angle projection remained approximately the same for both data sets, which also occurred with the other residual statistics.

Table 3 shows the estimated IOPs and estimated standard deviations with observations only in the hemispherical FOV for sensors 1 and 2. Table 4 presents the estimated IOPs and standard deviations for both Ricoh Theta sensors with the hyperhemispheric data set. The principal distance $c$ and principal-point coordinates ($x_0$, $y_0$) are presented in millimeters, whereas estimated standard deviations are presented in pixel units for easier analysis.

The estimated standard deviations of $c$, $x_0$, and $y_0$ were lower than 1 pixel for both models, as recommended for CRP applications (Campos et al. 2015). Indeed, the estimated standard deviations of IOPs obtained with the hyp data set (Table 4) were smaller than those estimated with the hem data set (Table 3), mainly for the equisolid-angle model. Comparing the distortion parameters, the equisolid-angle model presented lower values, which indicates that this projection better fits the hyperhemispherical observations for this camera. Another improvement from using hyperhemispheric observations can be noticed in the estimated standard deviations of EOPs, which can be interpreted as a measure of network quality.
Table 3. Estimated interior orientation parameters and standard deviations in the simultaneous camera calibration for the equisolid-angle and stereographic models with the hem data set.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Equisolid-Angle</th>
<th>Stereographic</th>
</tr>
</thead>
<tbody>
<tr>
<td>c (mm ± pixel)</td>
<td>1.4291 ± 0.18</td>
<td>1.4279 ± 0.22</td>
</tr>
<tr>
<td>x₀ (mm ± pixel)</td>
<td>0.0017 ± 0.14</td>
<td>0.0037 ± 0.14</td>
</tr>
<tr>
<td>y₀ (mm ± pixel)</td>
<td>0.3018 ± 0.18</td>
<td>0.2935 ± 0.20</td>
</tr>
<tr>
<td>K₁ (mm⁻¹)</td>
<td>4.78 × 10⁻⁵ ± 6.02 × 10⁻⁴</td>
<td>7.09 × 10⁻⁵ ± 8.57 × 10⁻⁴</td>
</tr>
<tr>
<td>K₂ (mm⁻¹)</td>
<td>6.06 × 10⁻³ ± 2.81 × 10⁻⁴</td>
<td>4.95 × 10⁻³ ± 3.94 × 10⁻⁴</td>
</tr>
<tr>
<td>K₃ (mm⁻¹)</td>
<td>−1.06 × 10⁻⁷ ± 4.02 × 10⁻⁸</td>
<td>−9.02 × 10⁻⁷ ± 5.56 × 10⁻⁸</td>
</tr>
<tr>
<td>P₀ (mm⁻¹)</td>
<td>2.03 × 10⁻⁵ ± 4.09 × 10⁻⁶</td>
<td>5.69 × 10⁻⁵ ± 5.09 × 10⁻⁶</td>
</tr>
<tr>
<td>P₁ (mm⁻¹)</td>
<td>5.06 × 10⁻⁴ ± 4.65 × 10⁻⁵</td>
<td>−1.12 × 10⁻⁴ ± 5.82 × 10⁻⁵</td>
</tr>
</tbody>
</table>

Table 4. Estimated interior orientation parameters and standard deviations in the simultaneous camera calibration for the equisolid-angle and stereographic models with the hyp data set.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Equisolid-angle</th>
<th>Stereographic</th>
</tr>
</thead>
<tbody>
<tr>
<td>c (mm ± pixel)</td>
<td>1.4309 ± 0.12</td>
<td>1.4319 ± 0.16</td>
</tr>
<tr>
<td>x₀ (mm ± pixel)</td>
<td>0.0032 ± 0.10</td>
<td>−0.0032 ± 0.12</td>
</tr>
<tr>
<td>y₀ (mm ± pixel)</td>
<td>0.3024 ± 0.10</td>
<td>0.2912 ± 0.14</td>
</tr>
<tr>
<td>K₁ (mm⁻¹)</td>
<td>7.58 × 10⁻⁷ ± 3.74 × 10⁻⁶</td>
<td>1.40 × 10⁻⁶ ± 5.23 × 10⁻⁶</td>
</tr>
<tr>
<td>K₂ (mm⁻¹)</td>
<td>8.06 × 10⁻⁶ ± 1.44 × 10⁻⁵</td>
<td>8.28 × 10⁻⁶ ± 1.89 × 10⁻⁵</td>
</tr>
<tr>
<td>K₃ (mm⁻¹)</td>
<td>−1.33 × 10⁻⁴ ± 1.71 × 10⁻⁵</td>
<td>−1.34 × 10⁻⁴ ± 2.13 × 10⁻⁵</td>
</tr>
<tr>
<td>P₀ (mm⁻¹)</td>
<td>2.10 × 10⁻⁴ ± 2.87 × 10⁻⁵</td>
<td>5.68 × 10⁻⁵ ± 3.65 × 10⁻⁵</td>
</tr>
<tr>
<td>P₁ (mm⁻¹)</td>
<td>6.57 × 10⁻⁴ ± 2.76 × 10⁻⁵</td>
<td>−8.81 × 10⁻⁵ ± 3.78 × 10⁻⁵</td>
</tr>
</tbody>
</table>

Table 5. Average relative orientation parameters and standard deviations calculated for the equisolid-angle and stereographic projection models with the hem and hyp data sets.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>hem</th>
<th>Equisolid-angle</th>
<th>Stereographic</th>
<th>hyp</th>
<th>Equisolid-angle</th>
<th>Stereographic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Δω₀</td>
<td>0.0181 ± 0.031°</td>
<td>0.0220° ± 0.152°</td>
<td>0.0027° ± 0.026°</td>
<td>0.0866° ± 0.332°</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Δφ</td>
<td>0.1129° ± 0.028°</td>
<td>0.0794° ± 0.075°</td>
<td>0.1047° ± 0.020°</td>
<td>0.0884° ± 0.098°</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Δx</td>
<td>−179.66° ± 2.721°</td>
<td>−179.65° ± 0.785°</td>
<td>−179.66° ± 0.782°</td>
<td>−179.66° ± 0.324°</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D₉ (cm)</td>
<td>1.89 ± 0.03</td>
<td>2.41 ± 0.03</td>
<td>1.95 ± 0.04</td>
<td>2.61 ± 0.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B₁ (cm)</td>
<td>0.067 ± 0.023</td>
<td>−0.053 ± 0.025</td>
<td>−0.074 ± 0.024</td>
<td>−0.472 ± 0.043</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B₂ (cm)</td>
<td>0.009 ± 0.042</td>
<td>−0.07 ± 0.038</td>
<td>−0.047 ± 0.049</td>
<td>−0.12 ± 0.091</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B₃ (cm)</td>
<td>−1.890 ± 0.035</td>
<td>−2.412 ± 0.039</td>
<td>−1.953 ± 0.041</td>
<td>−2.563 ± 0.059</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 5a shows the average values of the estimated standard deviations for camera orientation angles (σω, σφ, σx), and Figure 5b the average values of the standard deviations for camera positions (σx₀, σy₀, σz₀).

The estimated values of σω, σφ, and σx decreased by approximately 40% for the equisolid-angle model and 24% for the stereographic model when including the observations beyond a 180° FOV. The same result occurred with the standard deviations for the camera positions, except for the X-coordinate, which remained the same. Values of σy₀ and σz₀ had an average decrease of 24% with the equisolid-angle model and 9% with the stereographic model. After the simultaneous calibration of the Ricoh Theta S dual-fish-eye system, ROPs were computed from the estimated EOPs. Table 5 presents average values and standard deviations of ROPs (Δω₀, Δφ, Δx, D₉, B₁, B₂, B₃) obtained with the equisolid-angle and stereographic models for both data sets.

Higher standard deviations for Δω₀ and Δφ can be noticed with the hyp data set (Table 5). However, the standard deviations of the relative orientation angles (Δω₀, Δφ, Δx) were still lower than 1°, which is much better than the estimated standard deviations of the rotation angles (Figure 5). The use of hyperhemispherical points reduced the standard deviations of Δx and the base elements (B₁, B₂, B₃), with the equisolid-angle model achieving the best results.

The accuracies achieved in these trials were assessed with eight distances measured in the 360° calibration field. The RMSEs of the discrepancies between the reference and estimated distances for the equisolid-angle model were 6.81 mm with the hem data set and 6.53 mm with the hyp data set. For the stereographic projection model, the RMSEs were 6.76 and 6.34 mm with the hem and hyp data sets, respectively. The RMSEs of the discrepancies in control distances were compatible with the standard deviation of control points (5 mm) and the pixel size in object space units (approximately 5 mm). The RMSE values for both models were lower when adding observations in the hyperhemispherical image field, showing improvements in the simultaneous camera calibration of the Ricoh Theta S dual-fish-eye system.

Considering both fish-eye models used in Experiment B, the simultaneous bundle adjustment based on the equisolid-angle model presented better results than that based on the stereographic model. In this regard, the correlation coefficients between EOPs (c, x₀, y₀) and EOPs (α, φ, κ) were computed only for the equisolid-angle projection. Five images (10, 12, 20, 17, and 3) with different orientations were chosen for this analysis. Figure 6 presents the correlation coefficients between principal distance c and EOPs, showing camera orientation (α, φ, κ) and camera position (x₀, y₀, z₀) parameters. Figure 7 shows the correlation between the X-coordinate x₀ of the principal point and the camera orientation and position. Figure 8 presents the correlation between the y-coordinate y₀ of the principal point and the camera orientation and position. The correlation between estimated principal distance and camera coordinates (Figure 6a) presented a higher value with X₀, because the Ricoh Theta S camera was aligned with the x-axis of the reference system when taking these images, and thus X corresponds to depth. Due to this orientation, x₀ and y₀ had a higher correlation with Y₀ and Z₀, respectively (Figures 7 and 8). Image 17 presented an inverse correlation for x₀ and...
Figure 5. Average values of estimated standard deviation for (a) camera orientation angles ($\sigma_\omega$, $\sigma_\phi$, $\sigma_\kappa$) and (b) camera position ($\sigma_X^0$, $\sigma_Y^0$, $\sigma_Z^0$) for the equisolid-angle and stereographic models with the hem and hyp data sets.

Figure 6. Correlations between the principal distance $c$ and (a) camera orientation ($\omega$, $\phi$, $\kappa$) and (b) camera perspective center ($X^0$, $Y^0$, $Z^0$).

Figure 7. Correlation coefficients between the x-coordinate $x^0$ of the principal point and (a) camera orientation ($\omega$, $\phi$, $\kappa$) and (b) camera perspective center ($X^0$, $Y^0$, $Z^0$).

Figure 8. Correlation coefficients between the y-coordinate $y^0$ of the principal point and (a) camera orientation ($\omega$, $\phi$, $\kappa$) and (b) camera perspective center ($X^0$, $Y^0$, $Z^0$).
model the observations beyond a 180° FOV. The equisolid-angle and stereographic models presented better results for modeling hyperhemispherical lenses and can be considered suitable for the camera calibration of a dual-fish-eye system.

The use of hyperhemispheric points provided a better distribution of image points, and the degrees of freedom for bundle adjustment increased. Consequently, the estimation of calibration parameters improved, which indicates enhanced

![Figure 9](image.png)

**Figure 9.** Difference vectors of the targets in (a) 3D isometric view and (b) front view (X, Y) of the 360° calibration field. The scale of the residuals is 50 times the plot scale.

Table 6. Average relative orientation parameters and standard deviations calculated for the equisolid-angle and stereographic projection models with the hem and hyp data sets with and without stability constraints.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Equisolid-angle</th>
<th>Stereographic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Without stability constraints</td>
<td>With stability constraints</td>
</tr>
<tr>
<td>Δα</td>
<td>179.9776° ± 13.52°</td>
<td>179.9773° ± 0.152°</td>
</tr>
<tr>
<td>Δρ</td>
<td>0.0873° ± 1°</td>
<td>0.0795° ± 0.075°</td>
</tr>
<tr>
<td>Δκ</td>
<td>179.6599° ± 26.58°</td>
<td>179.6596° ± 0.785°</td>
</tr>
<tr>
<td>Dᵥ (cm)</td>
<td>2.398 ± 0.072</td>
<td>2.414 ± 0.038</td>
</tr>
<tr>
<td>Bᵥ (cm)</td>
<td>-0.087 ± 0.043</td>
<td>-0.053 ± 0.025</td>
</tr>
<tr>
<td>Bᵥ (cm)</td>
<td>-0.033 ± 0.073</td>
<td>-0.070 ± 0.038</td>
</tr>
<tr>
<td>Bᵥ (cm)</td>
<td>-2.396 ± 0.073</td>
<td>-2.412 ± 0.039</td>
</tr>
</tbody>
</table>

**Conclusion**

A study on rigorous fish-eye models (equidistant, equisolid-angle, stereographic, and orthogonal) for hyperhemispherical lens modeling and the use of stability constraints of ROPs for dual-fish-eye systems was presented in this article. In summary, the fish-eye models presented similar results when only points in the hemispherical FOV were used. Problems appeared with the addition of hyperhemispheric points, when the equidistant and orthogonal projections did not correctly
network geometry. Another improvement was observed in the correlation coefficient between TOPs and OOPs, which is an important issue for camera calibration. The best result for the Ricoh Theta S dual-camera calibration was achieved by the equisolid-angle model combined with the stability constraints of OOPs. The accuracy of the bundle adjustment was approximately the size of a pixel in object space units.

A slight increase in the observation residuals after bundle adjustment was observed with the addition of hyperhemispheric points. This can be explained by the nonuniform spatial resolution, which decreases in the border of fish-eye images. A discrete stochastic model to weight the observation as a function of the radial distance of the image point can minimize this effect, which is recommended for future works. Furthermore, other methods for the estimation of the covariance matrix can be investigated, such as variance component estimation.

In conclusion, the hyperhemispheric points, combined with a proper mathematical model, are beneficial for camera calibration. This provides a new perspective on the use of hyperhemispheric points for high-accuracy CRP applications. Further studies are required to evaluate the influence of hyperhemispheric points in other photogrammetric processes, such as 3D reconstruction.

Acknowledgments
This study was funded by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brazil, Finance Code 001 (grants 1774590, 88881.310314/2018-01, and 88882.433941/2019-01), and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (grants 404379/2016-8, 155739/2018-2, and 141550/2020-1).

References


Schenk, T. 2005. *Introduction to Photogrammetry*. Columbus, OH: Department of Civil and Environmental Engineering and Geodetic Science, The Ohio State University.


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

1. Introduction to DEMs  
   David F. Maune, Hans Karl Heidemann, Stephen M. Kopp, and Clayton A. Crawford
2. Vertical Datums  
   Dru Smith
3. Standards, Guidelines & Specifications  
   David F. Maune
4. The National Elevation Dataset (NED)  
   Dean B. Gesch, Gayla A. Evans, Michael J. Oimoen, and Samantha T. Arundel
5. The 3D Elevation Program (3DEP)  
   Jason M. Stoker, Vicki Lukas, Allyson L. Jason, Diane F. Eldridge, and Larry J. Sugarbaker
6. Photogrammetry  
   J. Chris McGlone and Scott Arko
7. IfSAR  
   Scott Hensley and Lorraine Tighe
8. Airborne Topographic Lidar  
   Amar Nayegandhi and Joshua Nimetz
9. Lidar Data Processing  
   Joshua M. Novac
10. Airborne Lidar Bathymetry  
    Jennifer Wozencraft and Amar Nayegandhi
11. Sonar  
    Guy T. Noll and Douglas Lockhart
12. Enabling Technologies  
    Bruno M. Scherzinger, Joseph J. Hutton, and Mohamed M.R. Mostafa
13. DEM User Applications  
    David F. Maune
14. DEM User Requirements & Benefits  
    David F. Maune
15. Quality Assessment of Elevation Data  
    Jennifer Novac
       Appendix A. Acronyms
       Appendix B. Definitions
       Appendix C. Sample Datasets

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

The NEEA documented the highest Return on Investment from QL2 lidar for the conterminous states, Hawaii and U.S. territories, and QL5 IfSAR for Alaska.

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

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

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

PRICING
Student (must submit copy of Student ID) $50 +S&H
ASPRS Member $80 +S&H
Non-member $100 +S&H
E-Book (only available in the Amazon Kindle store) $85
LEARN DO GIVE BELONG

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

asprs.org