ASPRS MEMBER EXCLUSIVE 20% DISCOUNT ON ALL WHITTLIES PUBLISHING BOOKS USING CODE WPASPRS2

WHITTLIES PUBLISHING'S STABLE OF CLASSIC GEOMATICS BOOKS INCLUDES THE THREE WINNERS OF THE PRESTIGIOUS KARL KRAUS MEDAL AWARDED BY ISPRS

- OBJECT AND PATTERN RECOGNITION IN REMOTE SENSING
  - Stefan Hinz, Andreas Braun and Martin Winharm
  - 978-184995-128-9

- 3D/4D CITY MODELLING
  - (Available as a CD)
  - 978-184995-475-4

- HIGH RESOLUTION OPTICAL SATELLITE IMAGERY
  - Second edition
  - 978-184995-390-0

- PRINCIPLES OF GEOSPATIAL SURVEYING
  - A. L. Allen
  - 978-184995-21-0

- REVIVING PALMYRA IN MULTIPLE DIMENSIONS
  - Images, Ruins and Cultural Memory
  - 978-184995-296-5

OUR LIST CONTINUES TO EXPAND WITH THESE NEW TITLES

BROWSE OUR WEBSITE TO SEE OUR FULL RANGE OF ACCLAIMED GEOMATICS BOOKS

WWW.WHITTLIESPUBLISHING.COM
Teledyne announces the launch of a geospatial group with the unification of its Optech and CARIS businesses. The new Teledyne Geospatial group will offer holistic solutions to seamlessly map land and sea through the integration of industry-leading lidar sensors combined with world-renowned software workflows.

Teledyne Optech has been a world leader in the design, development and manufacture of advanced lidar instruments for more than 45 years. And for over four decades, Teledyne CARIS has developed market-dominating software for the marine GIS community.

This collaboration provides customers with a competitive edge in mapping and delivering data products inside of one complete workflow. Ease of collection and processing, through to final product, is enhanced with efficiency-driving AI algorithms and real time quality control. Ongoing collaborations with other Teledyne businesses extends the geospatial capabilities even further.

The CZMIL SuperNova, the first product from Teledyne Geospatial, integrates Optech’s bathymetric lidar with CARIS’ comprehensive processing software providing the highest performing bathymetric lidar system in the world.

Andy Hoggarth, Vice President, Sales and Marketing at Teledyne Geospatial explains: “Businesses today are increasingly expected to offer a more comprehensive suite of services. Bringing Optech and CARIS together allows us to leverage the world-leading expertise of both companies, ensuring customers can fully realize the competitive advantages provided by our ability to deliver holistic solutions for land and sea.”

Teledyne Geospatial unifies the hardware and software expertise of two unique companies in the geospatial arena. The new group provides customers with innovative integrated solutions. Offerings will include turnkey systems, lidar and sonar integrated workflows and a range of systems and solutions that support holistic, precision data collection.

UP42, a geospatial developer platform and marketplace wholly owned by Airbus, has partnered with Indian startup HyperVerge Inc. to offer satellite imagery services using Artificial Intelligence (AI)-based change detection algorithms.

This is the 11th successful partnership industrialized by Airbus Bizlab in India and the second partnership between HyperVerge and an Airbus entity.

HyperVerge’s automated algorithms will help detect small structural changes to homes and properties for local government tax assessment and code enforcement. They will also help monitor legal and unauthorised construction of roads, buildings and artificial islands in remote regions by federal or defence agencies, and keep GIS maps updated in rapidly growing urban areas. The automated algorithms will detect these changes with a high degree of accuracy by comparing two images acquired on different dates.

“UP42 is excited to launch this partnership with HyperVerge, a company at the forefront of innovation around deep learning and satellite imagery,” said UP42 CEO Sean Wiid. “The sheer volume of satellite imagery makes manual interpretation difficult or even impossible. Algorithms like those developed by HyperVerge that automatically and accurately detect change are of utmost importance for applications ranging from infrastructure monitoring to urban planning.”

Users of the UP42 platform will gain affordable access to SPOT and Pléiades archives where they can purchase the image data required to cover their area of interest. Leveraging the power of cloud computing, users can then apply HyperVerge algorithms that join more than 130 data and analytics capabilities, to find features, count objects, detect change, uncover patterns, classify land use, and derive vegetative indices.

“HyperVerge is very excited for this partnership with UP42, which is changing the way satellite imagery-based applications are built and consumed,” said HyperVerge CEO Kedar Kulkarni. “Combining the ease of use of the UP42 platform with HyperVerge’s AI algorithms can help enterprises, defense agencies, and governments track changes in assets of interest with a high accuracy in tactically relevant timelines.”

Auterion, the company building an open and software-defined future for enterprise drone fleets, today announced that it has partnered with Phase One. The companies will make the Phase One P3 Payload line up easily accessible, with a plug-and-play integration to Auterion’s open drone ecosystem.

Drones leveraging the Phase One P3 Payload and the power and connectivity of Auterion’s Skynode and Suite are capable of dramatically scaling high-value, high-risk and time-critical inspections including those of wind turbine fields (on land and offshore), oil refineries and offshore rigs, power masts and utility lines, bridges, dams, nuclear facilities, large infrastructure projects and other use cases. The combination also benefits faster geospatial mapping, bringing world-renowned image quality with very high resolution, dynamic range, color fidelity and geometric accuracy to projects.

“The partnership between Auterion and Phase One makes it possible to scale drone inspection operations,” said Thomas Gubler, head of avionics at Auterion. “Customers will benefit from cloud connectivity, which shortens and simplifies workflow and enables faster corrective actions for any problem identified. There’s no more pulling from an SD card. A cloud-connected workflow becomes increasingly important when paired with the capabilities of Phase One’s high-precision, high-quality P3 Payload.”

A growing open ecosystem for drones provides an increas-
ing level of choice for customers. “More end users can leverage payloads for plug-and-play operation without extensive integration, which means enterprises can expand and diversify their fleets fast,” said Gubler. “Auterion’s easy, open standards software integration is yielding the most versatile and innovative range of drone solutions for scaled enterprise applications, like this uniquely powerful pairing with Phase One. The Auterion ecosystem future-proofs our offerings, de-risking investments in the eye of our customers.”

Learn more at www.phaseone.com or www.auterion.com.

**ACCOMPLISHMENTS**

URISA is pleased to announce that Josiah Burkett and Samantha Strang are the recipients of 2021 Dr. Marilyn O’Hara Ruiz Young Professional Scholarships. This scholarship fund was established in 2018 and honors Dr. Marilyn Ruiz. During Dr. Ruiz’s career at the University of Illinois at Urbana-Champaign, she provided her undergraduate, graduate and post-doctoral students with excellent research experiences in her laboratory which helped them to have successful careers in academia, government, and industry. Marilyn was passionate about her role as a mentor of graduate education.

URISA established the Dr. Marilyn O’Hara Ruiz Young Professional Scholarship program which selects up to two young professionals (35 years old or younger) to attend GIS-Pro in-person. The scholarship application is rigorous and the evaluation committee appreciates the effort and detail with which all candidates presented their qualifications.

Josiah Burkett is a GIS Analyst for GeoTechVision in Kingston, Jamaica. He is an active member of URISA’s Vanguard Cabinet and serves as a volunteer on URISA’s Community Resilience Committee and the Climate Change & Climate Equity Working Group.

Samantha Strang is a GIS Technician for Chatham County in Savannah, Georgia. Her knowledge of coastal dynamics and processes along with knowledge of fundamentals of elevation, along with her ability to fly drones with her FAA 107 license and airspace awareness, brings an added dimension to GIS offerings at Chatham County.

Fik Winata, one of the members of the scholarship review committee, expressed the following: “Both of the scholarship winners are undeniably exceptional. The committee was impressed with their contributions related to climate change, disaster management and recovery, and the coastal environment. Their efforts, within and outside of their jobs, have positive impacts on people’s lives and make a difference.”

URISA is pleased to support these two young professionals, along with so many others who are the future leaders of the organization. Learn more about the scholarship, and consider donating to support future winners, here: www.urisa.org/yp-scholarship.

**CALENDAR**

- 27-28 October, 7th International Conference on Engineering and Emerging Technologies, Istanbul, Turkey. For more information, visit www.iceet.net.
- 3-5 November, Survey & GIS Summit. For more information, visit https://surveygissummit.com/.
- 8-12 November, URISA GIS Leadership Academy, St. Petersburg, Florida. For more information, visit www.urisa.org/education-events/urisa-gis-leadership-academy/.
- 14-18 December, 30th International Cartographic Conference & International Cartographic Exhibition, Florence, Italy. For more information, visit https://icaci.org/icc2021.
- 1-3 February, 2022, URIS LEAP Conference. For more information, visit www.urisa.org/leap.
- 6-8 February 2022, Geo Week 2022, Denver, Colorado. For more information, visit www.geo-week.com/.

**NCASPRS Fall 2021 Conference**

Come to share, learn and network with the American Society of Photogrammetry and Remote Sensing in North Carolina

**Monday, November 15 – Tuesday, November 16, 2021**

In coordination with Fayetteville State’s GeoWeek

Presentation times are 1:00 to 5:00 PM on November 15th and 8:00 AM to 12 PM November 16th

The conference is virtual (via Zoom) with in-person* hybrid for workshop

*In-person location is tentatively scheduled at UNC Fayetteville

For more information or to register scan the QR Code
705 System Calibration Including Time Delay Estimation for GNSS/INS-Assisted Pushbroom Scanners Onboard UAV Platforms

Lisa M. LaForest, Tian Zhou, Seyyed Meghdad Hasheminasab, and Ayman Habib

Unmanned aerial vehicles (UAVs) equipped with imaging sensors and integrated global navigation satellite system/inertial navigation system (GNSS/INS) units are used for numerous applications. Deriving reliable 3D coordinates from such UAVs is contingent on accurate geometric calibration, which encompasses the estimation of mounting parameters and synchronization errors. Through a rigorous impact analysis of such systematic errors, this article proposes a direct approach for spatial and temporal calibration (estimating system parameters through a bundle adjustment procedure) of a GNSS/INS-assisted push-broom scanner onboard a UAV platform.

717 Least Squares Adjustment with a Rank-Deficient Weight Matrix and Its Applicability to Image/Lidar Data Processing

Radhika Ravi and Ayman Habib

This article proposes a solution to special least squares adjustment (LSA) models with a rank-deficient weight matrix, which are commonly encountered in geomatics. The two sources of rank deficiency in weight matrices are discussed: naturally occurring due to the inherent characteristics of LSA mathematical models and artificially induced to eliminate nuisance parameters from LSA estimation.

735 Spectral Reflectance Estimation of UAS Multispectral Imagery Using Satellite Cross-Calibration Method

Saket Gowravaram, Haiyang Chao, Andrew Molthan, Tiebiao Zhao, Pengchi Tian, Harold Flanagan, Lori Schultz, and Jordan Bell

This article introduces a satellite-based cross-calibration (SCC) method for spectral reflectance estimation of unmanned aircraft system (UAS) multispectral imagery. The SCC method provides a low-cost and feasible solution to convert high-resolution UAS images in digital numbers (DN) to reflectance when satellite data is available.

747 Early Classification Method for US Corn and Soybean by Incorporating MODIS-Estimated Phenological Data and Historical Classification Maps in Random-Forest Regression Algorithm

Toshihiro Sakamoto

An early crop classification method is functionally required in a near-real-time crop-yield prediction system, especially for upland crops. This article proposes methods to estimate the mixed-pixel ratio of corn, soybean, and other classes within a low-resolution MODIS pixel by coupling MODIS-derived crop phenology information and the past Cropland Data Layer in a random-forest regression algorithm.

759 A Deep Multi-Modal Learning Method and a New RGB-Depth Data Set for Building Roof Extraction

Mehdi Khoshboresh-Masouleh and Reza Shah-Hosseini

This article focuses on tackling the challenge of building mapping in multi-modal remote sensing data by proposing a novel, deep superpixel-wise convolutional neural network called DeepQuantized-Net, plus a new red, green, blue (RGB)-depth data set named I3D.

767 Feature-Point Matching for Aerial and Ground Images by Exploiting Line Segment-Based Local-Region Constraints

Min Chen, Tong Fang, Qing Zhu, Xuming Ge, Zhanhao Zhang, and Xin Zhang

In this article, we propose a feature-point matching method that is robust to viewpoint, scale, and illumination changes between aerial and ground images, to improve matching performance.
Covered with lakes, forests, and mountains, Dalarna County has been called “Sweden in miniature.” But the same region that today draws people to its idyllic lakeside villages and midsummer celebrations was also the site of an ancient, catastrophic impact.

Around 380 million years ago, in the Late Devonian period, an asteroid slammed into the land that is now south-central Sweden. The impact left quite a mark. Even after hundreds of millions of years of erosion, the scar is still recognizable. It is especially apparent when viewed from above.

The Siljan impact structure, or “Siljan Rings,” is visible in this image, acquired on June 24, 2020, with the Operational Land Imager (OLI) on Landsat 8. Measuring more than 50 kilometers (30 miles) across, Siljan is the largest-known impact structure in Europe and among the top 20 largest on Earth.

Surveys of the structure have shown that the ground is slightly raised up across parts of the crater’s center. It is surrounded by a ring-like graben, or depression, which today is partially filled with water. Lake Siljan, on the crater’s southwest side, is the largest lake; it connects to Lake Orsa via a small river.

People have lived for millennia near the crater without knowing its cosmic origin. In the late 1960s, scientists used drill cores to uncover the complex and ancient geology deep below the ground.

Research at Siljan is ongoing today. In a 2019 study, scientists described how they used drill cores to find that the deep, fractured rocks in the crater were suitable for ancient life. A subsequent paper in 2021 described the fossilized remains of fungi discovered at a depth of more than 500 meters.

Thermal Imagery for Building and Utilities Owners

By Woolpert’s Qassim Abdullah, Ph.D., PLS, CP and Nadja Turek, PE, F. SAME, LEED AP BD+C, GGP, Envision SP
**BACKGROUND**

**Thermography Defined**

Temperature, or heat energy, is a key variable in virtually all processes around us. In the human body, a deviation of even a few degrees from the normal body temperature of 98.6°F is an indication of infection. In the kitchen, baking a cake at the wrong temperature will prevent it from rising properly and result in poor taste and texture. Countless similar examples exist in facility and industrial applications.

Thermography, or the science and technique of conducting heat inspections using imaging sensors or cameras, visually captures minute changes in temperature and has many advantages over traditional contact-spot heat inspections. Among these advantages are:

1. It is non-contact in nature and keeps operators out of danger.
2. It is two-dimensional, which provides a fuller understanding for interpretation than a one-dimensional spot.
3. It can be conducted in real time and upon fast-moving targets.
4. It does not intrude upon examined objects.
5. It is non-destructive; for example, buried utilities can be inspected without excavation.

**Value of Thermal Imagery**

Imagine a scenario in which a water main, located next to a steam line, breaks and undermines the surrounding soil. The responding employee is badly burned by the heated water because the as-built map incorrectly located the steam line. An accurate thermal (i.e. heat) map of the area would have prevented harm to the employee.

Thermal imaging is a technology that captures differences in temperature and can be used to locate anything that emits heat—including utilities. Clients regularly report using thermal imagery to identify missing or incorrectly located utilities. Updated records with accurate thermal imagery locations enhance safety for employees and customers.

Recognized as an effective, non-destructive method for inspecting buried utilities, thermal imaging helps facility engineers identify potential problem areas early and prioritize planned maintenance efforts. Through this technique, it is possible to inspect utilities, such as steam and hot water, for heat loss without damaging the surrounding pavements. However, much of today’s current thermal imagery is low-resolution, geospatially inaccurate, and imprecise, leaving imagery interpretation and corrective action planning to the client.

Woolpert is pioneering high-resolution, geospatially accurate thermal imagery with both automated and engineering analysis to bring value to facility and utility managers.

Energy audits provide a systematic approach to discovering energy issues and making sound decisions. The primary goals of an energy audit are to qualify and quantify the performance of building energy systems, identify opportunities for improvement, and assess the potential outcomes of those improvements, both in financial and non-financial terms.

Time intensive and expensive, yet ripe for innovation, building energy audits require qualified auditors to acquire on-site data, one building at a time. However, sizable campuses such as military bases, utility service areas, business parks, or university campuses are often so large and diverse, with multiple building construction types and uses, that the traditional approach to energy auditing can be not only time but also cost prohibitive. For example, an ASHRAE Level 2 audit that runs, on average, $0.10 to $0.15 per square foot, could cost a single campus hundreds of thousands of dollars and take several years to complete, before capital improvement projects even begin.

Facility managers need actionable energy audits in months, not years, to save money, energy, and emissions. Innovative auditing methods and cutting-edge technology are necessary for transforming the auditing industry for large facilities, campuses, and land holders.

Woolpert invested research and development resources into exploring the use of digital thermal imagery for inspecting underground utilities and analyzing the energy loss of residential and industrial facilities. A team of researchers from Woolpert and the University of Dayton developed and tested a new remote sensing-based method for conducting information-based energy audits. The long-term objective was to use thermal imagery to facilitate comprehensive energy audits and energy loss analysis with limited on-the-ground resources and reduced timeframes.

Remote auditing techniques enable the accurate identification of numerous energy efficiency measures (EEMs) without the necessity of building visits. This approach integrates readily accessible data—GIS building and property data, historical utility energy use data, historical weather data, and aerial building thermal imagery (thermography)—with advanced remote sensing technology to analyze the energy performance of a building without stepping foot on site.

Woolpert also developed and introduced an enhanced version of a regional heat score map (HSM) as an intuitive way to visualize energy efficiency problems and motivate owners or utilities to implement EEMs. Originally developed by the University of Calgary, Canada, the HSM is a GIS-based map that uses thermal imagery and modeling algorithms to evaluate energy loss from building rooftops. This visual tool compares heat loss across buildings in a specified region, enabling owners/managers to easily identify the roof area(s) needing inspection and/or repair. When offered by utility companies, the tool helps customers to view their heat scores and easily contact the utilities or recommended contractors to initiate needed improvements. The tool also has the potential to be used by private or government owners of large building inventories to manage energy consumption.
Methods of Collection

The preferred mode of thermal collection is largely dependent on the size of the collection area and the goal of the project. Manned aircraft, unmanned aircraft systems (UAS), and handheld scanners are all methods that, correctly tailored and applied, can be used for efficient and cost-effective thermal imagery collection.

Manned Aircraft

For large projects requiring an overall campus collection, manned aircraft mounted with thermal cameras and natural-color digital cameras are often the best collection tools. Manned aircraft can remain in the air significantly longer than most UAS, and they can cover large footprints with speed and efficiency. In addition, manned aircraft can acquire both nadir and oblique imagery to cover the roof and four building facades.

UAS

UAS is often impractical for large-collection flight plans. Limited by battery life and extreme temperature conditions, UAS typically have much shorter flight times than manned aircraft. Additionally, UAS flight speed is significantly less than that of the manned aircraft used to collect aerial imagery. However, for smaller flight footprints involving no more than a handful of buildings, UAS can offer an efficient, cost-effective solution.

Mobilizing UAS is significantly easier and less expensive than mobilizing manned aircraft for small-scale projects. Preferred over handheld scanners when entire building envelopes require assessment, UAS can collect imagery of roofs, difficult-to-access building faces, and exterior building walls.

Current Federal Aviation Administration regulations concerning UAS do present a major challenge to using UAS for thermal imagery acquisition. Even with the latest and less-rigid regulations represented by Part 107, UAS are not permitted to fly over people not involved with the operation, nor can they be flown at night without difficult-to-obtain waivers from the FAA. These restrictions make it difficult to employ UAS for projects located within inhabited areas and most likely need to be flown at night.

Handheld Scanners

Handheld thermal scanners or cameras are typically the most cost-effective tool for most indoor applications, and they also may be the preferred tool for outdoor applications when only a few buildings require scanning and roof imagery is not required. Using a handheld scanner requires no more mobilization than traveling to the job site with the scanner.

Image Collection and Processing

Thermal cameras with radiometric sensors are calibrated to measure the absolute temperature of every pixel in the image. Woolpert uses a state-of-the-art thermal sensing technology, the FLIR SC8300 HD science-grade cooled sensor (Figure 1). Designed with cutting-edge functionality for scientists and researchers, the SC8300 HD is a high-speed, mid-wave infrared (MWIR) camera with highly sensitive cooled InSb detectors and superb resolution. Woolpert interfaces the SC8300 with an airborne GPS system to capture the exact position in space of each thermal image during the mission. The entire system is mounted on a Woolpert aircraft modified for aerial imagery acquisition missions. The thermal imagery is usually acquired after midnight from an altitude of approximately 3,000 feet above ground level to produce thermal imagery with a ground sampling distance of one foot.

The acquired imagery provides thermal values for each pixel in a raster. Once corrected for emissivity and bias, these values represent individual temperatures for every pixel. The imagery is then produced in 32-bit format and processed following a stringent photogrammetric workflow to create a seamless mosaic with pixel-specific temperatures.

Complicating Factors

Our research has highlighted multiple variables impacting the quality and accuracy of the data collection.

While the camera captures return energy emitted by a surface, factors including material type and age, reflected temperature, weather conditions, and clouds between the aircraft and surface can limit the ability to obtain precise temperature values.

These variables dictate the need for robust ground-truthing to correct for atmospheric conditions and other energy-scattering phenomena. Woolpert uses handheld thermal cameras, the FLIR TGI165 and FLIR C5 (Figure 2), to measure...
and document the temperature of different ground surfaces during aerial thermal imagery acquisition. To ensure positional accuracy of the thermal mosaic, a set of ground control points is surveyed and used for aerial triangulation of the imagery.

**Imagery Collection Considerations Logistics**

Once the appropriate mode of collection has been determined, the conditions of the collection must be analyzed. The flight plan for an aerial acquisition mission is mainly dictated by desired quality, required accuracy, and sensor geometry. Once a suitable GSD is chosen for the project, then the flying altitude, sensor lens focal length, and sensor array size (i.e., how many charge coupled devices (CCD) make up the sensor) can be determined.

We recommend higher ground resolution thermal imagery (i.e., 15-centimeter or 30-centimeter GSD) be flown from lower altitudes to provide the most accurate thermal modeling. Flying low places the thermal camera closer to the structures under investigation and shortens the atmospheric column that the emitted thermal energy must pass through to register with the sensor.

In addition to the thermal imagery, airborne GPS data should also be collected to support the processing and aerial triangulation required to create high-resolution, geospatially accurate images. If there is no existing colored aerial imagery of the area, natural-color imagery may also need to be flown during the day to support object verification and analysis.

The next step in collection planning is to decide if nadir (directly overhead) or oblique (at an angle) imagery is needed. The study target and objective clearly impact this decision. If assessing only roofs or utilities, then nadir collections may suffice. Nadir imagery can be collected with simple, straight flight paths back and forth over a collection area. However, if thermal imagery of exterior walls or complex building geometries is desired, then oblique collections will be necessary. In oblique collections, the sensor is aimed at off-nadir angles, usually 45 degrees, to provide side views of buildings and objects (see Figure 3). Oblique imagery collections are more time consuming and expensive than nadir collections because the flight path must cross the building multiple times in multiple directions to image every side of the building.

Finally, the following logistics considerations contribute to the complexity of image collection planning:

- Controlled airspace
- Security clearances requirements (both flight and ground personnel)
- Ground control network needs
- Ground truthing or field measurement recording needs

**Nature**

Atmospheric conditions also impact the quality of a collection effort. Because large temperature differentials appear more vivid in thermal imagery, winter is the preferred season for collecting thermal imagery. On hot summer days, there may be only a 15- or 20-degree difference between indoor and outdoor temperatures; however, depending upon location, cold winter nights can easily provide more than 60 degrees in temperature difference.

Thermal imagery is typically collected at night to reduce, as much as possible, the impact of thermal clutter, solar flux, and reflection. It is best to wait several hours after sunset to reduce confusion between objects heated by daytime solar energy and problem hot spots caused by energy leakage. Fog, rain, snow, or wind can skew thermal image appearances, so clear, dry nights are recommended for providing the best contrast, and ultimately, identifying envelope issues.

**Sensor Resolution**

The thermal camera itself has design characteristics that affect the ground resolution of the resulting thermal imagery. The detector geometrical resolution (i.e., the size of individual pixels and how many pixels are in the detectors (array)) determines the granularity of the collected data. While high resolutions improve detail visibility, they are not always necessary and should be employed on a case-by-case basis.

**Sensor Sensitivity**

Thermal sensitivity indicates the smallest temperature difference between two objects that the camera can detect. Less-sensitive thermal cameras will provide values rounded to 0.2 degrees Celsius, while more-sensitive cameras may provide values to the 0.1 or even 0.01 degrees Celsius. While these differences may appear minimal, they significantly impact the crispness and granularity of the thermal imagery. Because thermal sensitivity directly affects the quality of the data collected, it naturally affects the analysis of that data.
Camera Settings
The three most important camera settings that must be verified prior to data collection are lens focus, temperature range, and proper optical distance. Thermal camera operators must understand the importance of adjusting the focusing ring of the camera to obtain clear, focused imagery. Good, consistent focus is critical to proper temperature measurements.

Temperature range is the minimum and maximum temperatures that the camera can measure and record. Once this range has been set for an imagery acquisition session, it cannot be changed at any point during data processing. Setting the correct temperature range on the sensor prior to data acquisition will result in optimal data quality because the thermal sensing ability will be optimized for the situation at hand. Expanding the range beyond that of the temperature of ground objects within the project boundary will reduce the sensor’s ability to discriminate between objects with subtle temperature differences.

Setting the appropriate working distance to the target is the last of the three settings that the operator needs to check before acquiring data. Acquiring thermal imagery from the proper distance to the target results in proper resolution of the imagery and therefore proper target details.

During data post processing, most settings can be adjusted to enhance the imagery quality; however, the three previously mentioned cannot. One extremely important setting that can be altered during processing is object emissivity. Emissivity measures how well materials absorb and emit thermal radiation. Every material has a unique emissivity value, defined as a ratio of how well it emits thermal radiation compared to an ideal black surface at the same temperature (which is assigned a value of 1.0). Emissivity directly impacts the temperature the thermal camera detects. Inflated emissivity will inflate the recorded temperatures, and vice versa. Unlike temperature range, emissivity can be adjusted during data export and processing. Because emissivity has such a significant impact on imagery accuracy, it will be discussed again in “Importance of Emissivity.”

![Figure 4. Buried pipes at the University of Dayton campus appear as bright (hot) linear features under roads and parking lots.](image)

Quantitative and Qualitative Results
Thermal imagery provides opportunities for both quantitative and qualitative analysis. It can be displayed either in color or grayscale, with each color or gradient representing a range of heat emission or temperatures. The temperature range associated with different colors is a user preference that varies from project to project, though blue is commonly used to represent cool temperatures and red is often used for hot temperatures.

Collecting accurate quantitative data is often challenging with thermal cameras because of the multiple variables already discussed, including emissivity, reflected temperature, atmospheric conditions, and atmospheric interference. Because of the uncertainty created by these variables, quantitative data is more frequently used for analysis.

Qualitative data enables comparative analysis between the heat signatures of similar objects for the purpose of identifying anomalies indicative of performance issues. Qualitative analysis is commonly used for inspection of roofs as well as buried and above-ground utility networks following the image interpretation concept that we use in remote sensing and mapping applications (Figure 4). While it is easier to perform, qualitative analysis alone doesn’t provide enough data for an energy audit, which requires metrics for temperature and heat transfer. A quantitative analytic approach is essential for energy auditing because it provides a method for quantifying energy and cost-reduction measures. It is important to understand the limitations of quantitative analysis and the controlled environment it requires budget for and mitigate (where possible) uncertainty and error.

Importance of Emissivity
Emissivity presents a variety of challenges to thermal imagery analysis. The limitations of determining emissivity values are a continuous hinderance, though emissivity tables can be useful for identifying approximate values for common materials. Additionally, because dissimilar materials have different emissivity values, five different materials in a single thermal image will present five different emissivity values. To obtain accurate temperature values, emissivity values must be associated correctly with their corresponding materials during image export, increasing the likelihood for multiple exports.

Another analysis challenge is that many materials do not have known emissivity values. In other cases, environmental conditions or material finishes may dramatically alter known emissivity values. A strategy for determining unknown values is to place materials of known emissivity and temperature within the same image frame to create reference points. This strategy is labor intensive but effective for identifying unknown emissivity values. To date, there are no handheld tools capable of providing accurate, real-time emissivity measurements.
**DATA ANALYSIS**

Once the imagery has been acquired with the appropriate platform, thermal camera, and applicable camera settings, the imagery must be exported with the appropriate emissivity. This exported imagery is then processed using Woolpert’s photogrammetric workflow to create seamless, georeferenced mosaics. By interpreting the qualitative mosaic or performing quantitative modeling, one can identify issues such as building envelope or moisture leaks, mechanical or electrical equipment problems, or thermal bridging concerns.

Often, envelope issues such as missing insulation, leaks, or thermal bridging are invisible to the naked eye. Thermal imaging reveals these envelope deficiencies as warmer relative to the surrounding envelope. When imagery is flown over a heated building on a cold evening, missing insulation and leaky window seams will present as warm areas on the thermal image. Thermal bridging also will appear warm on the thermal image. Thermal bridging occurs when more conductive materials (such as steel beams) provide less-resistant paths for heat transfer across thermal barriers (such as insulation).

Moisture issues are also picked up by the thermal camera for two reasons. The first reason is that evaporation cools material surfaces such that wet areas appear cooler than the surrounding envelope. Alternatively, if moisture is trapped in the roof, the roof may appear as a warmer area because the heat rising from the building will increase the temperature of the water vapor more quickly than the surrounding area. Because moisture can reduce the effectiveness of some kinds of insulation, such as batt insulation, areas of insulation damaged by moisture may allow even more heat transfer. The temperature difference of an area relative to the surrounding envelope, walls, roof, or windows can indicate envelope issues that never would have been revealed otherwise.

Another useful application of thermal imagery analysis involves examining mechanical and electrical equipment. Malfunctioning equipment often appears warmer than nearby equipment. Hot spots on equipment may indicate that an inordinate amount of energy is emitted as waste heat instead of as the electric energy needed to perform the intended task. Motors, belt drives, boilers, pumps, and compressors are mechanical components that can be inspected for performance using thermal imaging. Electrical equipment such as breakers, wires, and circuits can be inspected similarly. Electrical components operating warmer than surrounding elements may be malfunctioning and should be further inspected. Thermal imaging can offer an alternative look when manual inspection does not reveal any malfunction.

Ductwork and piping can be inspected for issues using thermal imaging. Depending whether the system is in heating or cooling mode, ductwork and piping leaks appear warmer or cooler than the surrounding areas. Deteriorated pipe insulation can be identified in a manner similar to that used for moisture-degraded envelope insulation. Condensate pipe leaks also may be detected with warm or cool spots (depending on condensate type). Piping for outside utilities, such as steam, hot water, or chilled water, can be inspected with thermal imaging. Both above- and below-ground piping will appear in the imagery, with large hot or cool spots indicating leaks or damaged insulation. Figure 4 illustrates the clear thermal signature, shown in bright linear feature, for the buried hot water and/or steam pipes at the University of Dayton campus. Thermal imaging provides facility engineers with an accurate, non-destructive method for inspecting underground networks.

The diversity of thermal imagery applications makes it useful for a wide range of services. Thermal imagery can aid in prioritizing renovations, such as envelope repair, roof replacement, and mechanical or electrical equipment rehabilitation. Aerial thermal imagery is especially useful for roof inspections and comparisons. Campus facilities often conduct building inspections to determine the order of building renovations, and thermal imagery can assist greatly with those building inspections. Thermal imagery also informs energy audits. Building envelopes have been difficult to include in energy audits due to variables such as the quality of insulation, thermal bridging, and moisture impacts. Thermal imagery enables energy audits to provide a more accurate picture of needed envelope upgrades. Finally, thermal imagery can be used to analyze utilities located outside of the building or beneath the ground and identify areas requiring repair or replacement.

**Heat Score Methodology**

To better understand the power of thermal imagery for comparative analysis, Woolpert used thermal imagery to develop an HSM for a residential neighborhood on the University of Dayton campus. The HSM is a GIS-based map that visually depicts the thermal efficiencies of building rooftops and compares them to neighborhood houses of similar construction.

After Woolpert collected the thermal imagery, the project team created a raster mosaic with temperature values stored in every pixel (Abdullah, et al. 2015). The set was corrected for variables, such as emissivity, that could introduce error to the values. Because rooftop temperatures can be skewed by chimneys (emitting heat) or tree branches (obscuring the imagery), the team identified these objects and removed their thermal pixels from the heat score calculation.

Heat scores are developed relative to the energy efficiency or heat loss of the surrounding buildings. Heat scores range from 0, the lowest possible waste heat compared to surrounding buildings, to 100, the highest possible heat loss compared to surrounding buildings. Using the heat score methodology developed by Bharanidharan Hemachandran at the University of Calgary (Hemachandran 2013), the heat score (based on the wasted heat modeled through the rooftop) is derived.
and attributed to the building polygon in the database. According to Hemachandran, a building’s estimated heat waste is calculated by multiplying the building’s living area by the difference between the building’s average rooftop temperature and the minimum temperature recorded in the study area. Areas of significant heat loss (i.e. hot spots) will appear on the HSM and indicate insufficient insulation or water damage. The Woolpert-developed HSM uses a modern, web-based portal to provide users with thermal images of their area and color-coded heat scores with pop-up boxes explaining how their houses rank compared to other houses in the area (in terms of heat loss). See figures 5, 6, and 7.

Each house in figures 5, 6, and 7 is assigned a color, either cool (blue or green) for low heat loss or warm (red or orange) for higher heat loss. The HSM helps users visualize how their homes currently are and should be performing compared to other neighborhood homes. Heat loss translates to higher utility costs, and owners made aware of high energy losses are more likely to make fixes because they can see for themselves the impact of poorly functioning building envelopes. More specific thermal imagery, such as individual rooftop images and hot spot highlights, can further help owners address heat loss problems.

**Case Study—University of Dayton**

**Challenge**
Partnering with researchers at University of Dayton, Woolpert investigated changes to the energy auditing process occurring if thermal imagery replaced traditional boots-on-the-ground auditing techniques.

**Approach**
Before any imagery was flown or analyzed, historical data over the previous year was analyzed to establish trends, baselines, and energy intensity for the various energy sources. As in a traditional energy audit, the team first conducted a utility bill analysis. Then, we used the historical data to perform regression analysis and determine the balance temperature for heating and cooling. Once weather and ground conditions were optimal, the team collected both nadir and oblique imagery to determine envelope values for the walls and the roof. The thermal imagery was collected at night to
minimize thermal clutter, solar flux, and reflection; it was paired with the plane’s onboard GPS to create high-resolution, geospatially accurate thermal datasets. Each building surface of the thermal dataset was isolated, and the minimum, maximum, and average temperatures were determined using the typical emissivity for metal roofing and siding. These temperatures were also calculated at higher emissivity values to simulate emissivity changes due to material aging/weathering. From an emissivity of 0.14 to 0.2, temperatures varied by roughly 4 to 5 degrees Fahrenheit. From an emissivity of 0.2 to 0.4, temperatures varied even more dramatically from 13 to 16 degrees Fahrenheit and demonstrated how reliant accurate temperature values are on correct emissivity values.

University of Dayton researchers developed a system of equations for determining the thermal resistance (R-value) of envelope components, walls, and roofs, based on the material’s exterior surface temperature. This process uses a dynamic model assuming an R-value to solve for a theoretical exterior temperature, compares it to the actual temperature, and then iterates until the error between the theoretical and measured temperature is near zero, this methodology is further explained in the paper “Estimating Envelope Thermal Characteristics from Single Point in Time Thermal Images” (Alshatshati 2018). Researchers inserted the Woolpert-collected temperature values into this model, and using various emissivity and corresponding temperatures, calculated R-values for the roofs and walls of a few buildings. The R-values ranged from 17.6 to 23.3 ft² F/Btu, depending on the emissivity used. Wall R-values saw less of a variation based on emissivity and ranged from 20.4 to 23.5 ft² F/Btu.

Outcome

Based on the analysis of the thermal imagery, it was determined that thermal imagery alone is not yet a viable and complete replacement for boots-on-the-ground energy audits. However, historical energy data coupled with thermal imagery can help prioritize the buildings for which boots-on-the-ground efforts will maximize the return on investment. Additionally, thermal imagery presents valuable potential for envelope analysis and identifying the roofs or walls with heat loss and insulation issues. As demonstrated, emissivity is crucial to providing accurate values for such an assessment, and as more data is collected on various material emissivity values, thermal imagery will become increasingly important for quantifying building envelope performance.

Conclusion

Thermal imagery is and has been an extremely useful tool for analyzing and inspecting building envelopes, above-ground and buried utilities, and mechanical and electrical equipment performance. It provides a layer of inspection of which the naked eye is not capable. Different collection vehicles add flexibility, offer cost-effective alternatives, and provide the ability to scale collections. Qualitative analysis has long been the preferred method of analysis, but as emissivity values are more accurately defined, quantitative analysis will become more viable and add increased capabilities. The ability to obtain accurate R-values enables auditors to quantify the savings of building envelope improvements, which have traditionally been difficult energy saving measures to calculate. At the community level, the HSM is an incredibly valuable tool that provides individuals with specific information about how their buildings are and should be performing within their environments. Overall, thermal imagery is a diverse tool that benefits individual homeowners and campus facilities personnel alike. For utility workers, thermal imagery can save lives by accurately locating underground utilities. As the accuracy of thermal imagery increases, its value and applications will grow in step.

References

Authors
Qassim Abdullah, Ph.D., PLS, CP, Woolpert Vice President and Chief Scientist has more than 40 years of combined industrial, research and development, and academic experience in analytical photogrammetry, digital remote sensing, and civil and surveying engineering. He is an ITC level II-certified thermographer. When he’s not presenting at geospatial conferences around the world, Dr. Abdullah teaches photogrammetry and remote sensing courses at the University of Maryland and Penn State, authors a monthly column for the ASPRS journal PE&RS, and mentors Woolpert’s research and development activities.

Nadja Turek, PE, F. SAME, LEED AP BD+C, GGP, Envision SP is a civil engineer and sustainable design expert serving as Woolpert’s sustainable design team leader. A former faculty member at the University of Dayton and the Air Force Institute of Technology, she engages in consulting, research, design and planning work. Ms. Turek works with teams of owners, design, and construction professions, using technology to evaluate sustainable and resilient design alternatives and equipment choices.
With ever increasing access to new datasets, we tend to include more datasets into the mapping application. When the Table of Contents fills up, we either start turning off layers or making groups. Eventually it becomes difficult to navigate to a specific item. So, here are a few quick tips and tricks to facilitate navigation in your map document.

A common feature of most, if not all, GIS software is a coupling between the geometry of a feature and the feature's record in an attribute table. That is, if you select an item's geometry on the map, it appears highlighted on the map and that same item is also highlighted in the layer's attribute table. Conversely, selecting an item in the attribute table, highlights the feature's geometry on the map. Again, like ArcGIS, most GIS software will allow the user to drive to a selected item, by double-clicking on the record in the attribute table.

To drive to the selected item, double-clicking on the highlighted record zooms-in to the item and centers the item in the map as in Figure 2.

But, what if you just want to center the selected item without changing the zoom ratio (map scale) to facilitate seeing a selected feature in a larger context, as in Figure 3?

There are actually two ways to accomplish this task. The conventional keystrokes are just to right-click on the item in the attribute table and select the “Pan To” option as in Figure 4. The hidden feature tip is to hold down the [CTRL] key and double-clicking on the gray-space to the left of the selected feature. Both methods will center the selected feature without changing the map scale.

There are lots of hidden gems in the ArcGIS user interface. Please feel free to share yours with us. Send your questions, comments, and tips to GISTT@ASPRS.org.

Kristopher Gallagher is a Senior LiDAR Analyst focused on the production of digital mapping products derived from topographic lidar and utility surveys. As senior geospatial scientist, Al Karlin works with all aspects of Lidar, remote sensing, photogrammetry, and GIS-related projects.

Photogrammetric Engineering & Remote Sensing
Vol. 87, No. 10, October 2021, pp. 697.
© 2021 American Society for Photogrammetry and Remote Sensing
doi: 10.14358/PERS.87.10.697
ANNOUNCING NEW STUDENT ADVISORY COUNCIL (SAC) COMMITTEE MEMBERS

We are delighted to welcome our new council members and a warm welcome back to our promoted councilors. All of our council members possess unique backgrounds and diverse experiences which makes them great assets to SAC and ASPRS.

Under the leadership of Evan Vega, Council Chair, and Lauren McKinney-Wise, Council Deputy Chair, this year’s SAC is planning to have a variety of events. We would also like to welcome Oscar Duran, a 2020 ASPRS scholarships winner, and Kenneth Ekpetere, who plans to work with NASA someday, as heads of their respective councils. Rabia Munsaf Khan, a Fulbright scholar and quintessential altruist, is our new Communications Council Chair.

All the members at ASPRS SAC strive to maximize student involvement so that they can benefit the most from the offered opportunities such as scholarships and webinars.

Since its inception in 2006, SAC has and will continue to offer amazing leadership opportunities for geospatial students through webinars and in person events. The prime focus of SAC is to build confidence among future geospatial leaders and encourage them to step outside of their comfort zones, learn new skills, and take advantage of opportunities provided by SAC and ASPRS. SAC is always looking for social media ambassadors, interested candidates can contact us at sac@asprs.org.

Lastly, we would like to thank our past Chair, Youssef Kaddoura, for successfully leading the team in achieving our 2020 goals. We truly value his involvement for this year and his unlimited support by providing insight, guidance, and dedicated service.
The republic is populated primarily by the Bantu group of peoples and is divided into two major language groups. The Shona speaking Mashona constitute 75 percent of the population and the Sindebele speaking Ndebele constitute about 20 percent of the population. The latter group arrived in the southwest around Bulawayo within the last 150 years and is an offshoot of the South African Zulu. They maintained control over the Mashona until the European occupation in 1890. Stoneage implements have been found in Zimbabwe, and ruins suggest an early civilization. The “Great Zimbabwe” ruins are located near Masvingo, and evidence suggests that it was built between the 9th and 13th centuries by Africans that established trading contacts with commercial centers on the continent’s southeastern coast. In 1888, Sir Cecil Rhodes obtained a concession for mineral rights from local chiefs and later the area was proclaimed a British sphere of influence. The British South Africa Company was chartered the following year and Salisbury (now Harare, the capital) was established in 1890. In 1895, the area was formally named Rhodesia in honor of Sir Cecil. The United Kingdom (PE&RS, October 2002) annexed Southern Rhodesia from the South Africa Company in 1923. A 1961 constitution was formulated that favored Caucasians in power. In 1965, the government declared Rhodesian independence, but the UK did not recognize the act and demanded more complete voting rights for the black majority in the country. United Nations sanctions and a guerilla uprising finally led to free elections in 1979 and independence as Zimbabwe in 1980. A land redistribution policy in 2000 has caused an exodus of white farmers, it has crippled the economy, and it has caused widespread shortages of basic commodities.

Zimbabwe is slightly larger than Montana, it is landlocked, and it is bordered by Botswana (813 km), South Africa (225 km), Zambia (797 km), and Moçambique (1,231 km). (PE&RS, September 1999). The terrain is mostly high plateau with a high veldt with mountains in the east. The lowest point is the junction of the Runde and Save Rivers at 162 m, and the highest point is Mount Inyangani (2,592 m).

In 1901, Alexander Simms completed a chain of quadrilaterals that spanned the west central part of the country. This chain started in the south near Bulawayo with the Inseza Base observed in 1898, it passed through Salisbury (now Harare) with the Gwibi Base observed in 1900, and it terminated in the north about 75 km east of where the Kariba Dam is now located on the Zambezi River. The geodetic coordinates of all the stations were referred to the origin point in Salisbury where: $\Phi_0 = 17° 50´ 25.440˝$ S and $\Lambda_0 = 31° 02´ 19.000˝$ E, with an azimuth to Mt. Hampden...

That article brought back some memories from long ago. In the late 1960s to early 1970s I was a commissioned officer (Topographic Engineer) at the U.S. Army Map Service (AMS) during the CORONA Program that utilized photographic products of military spy satellites. The systems used multiple star sensors to determine camera orientations similar to what is discussed by Sheng et al. Back then it was before the advent of digital imagery, so we used film that was physically returned from orbit. A special building was constructed at AMS to house the original film processing, duplicating, photogrammetric analyses, and cartographic stereocompilation of the final products.

Long ago, I was told a story about that building’s design by my civilian supervisor and mentor, Mr. Zeno V. Kittrell. Seems that the architect was informed that the building was going to house special film measurement instruments called “comparators” that were being custom-built to be capable of ultra-high precision and would be mounted on special massive granite bases. These instruments were termed “Semi-Automatic Coordinate Readers” (SACRs) and were intended to measure stellar camera images from the satellites. With a multitude of stars being imaged on each 70mm film frame, multiple cameras photographed per each cartographic frame; the task was expected to be particularly arduous. Hence the “semi-automatic” feature was to be of great advantage. With particular attention to detail, the building’s architect researched from which quarry the granite bases were to sourced, the weight of each finished base was computed, and the freight elevator load capacity for the building was then specified to be able to accommodate the SACR bases, one at a time. The room for the ultra-precise measurements of all imagery was “the Slab,” excavated a couple stories below ground, and “the slab” was a monolithic slab of concrete separate from the building and was “floating” on a thick cushion of neoprene. The “slab” had filtered laminar flow air conditioning, and even had a seismograph to maintain a record of the stability of the instrumentation.

As Zeno told me, the momentous day arrived that the SACRs were to be delivered by a caravan of flat-bed 18-wheel trucks. When the first truck started backing up to the loading dock, the steel doors electrically opened, a half-dozen Federal Officers came out on the loading dock with loaded 12-gauge riot guns to stand guard (I guess in case one of the drivers was a Russian spy), and a heavy-duty forklift truck came out of the building to accept the first SACR granite base. The forklift picked up the SACR from the flat-bed truck, backed up on the loading dock, and then proceeded into the building to the special freight elevator so that the base could be taken down into the sub-basement and “The Slab.” The forklift operator remained seated at the wheel, he reached up to the suspended elevator control box and had the safety doors close. He then pressed the “down” button, and the special freight elevator, the SACR base and the forklift truck instantly disappeared as the operator screamed as the elevator and all crashed down two floors. As Zeno related the catastrophe with tears running down his cheeks as he laughed, it seems the architect forgot to add the weight of the forklift truck to the freight elevator specifications! The rattled operator was given the rest of the day off, one of the guys in the machine shop ran the forklift truck for the Slab level, and another machinist got a second forklift to operate at the loading dock to load the freight elevator sans forklift. I’ll bet those granite bases are still on the Slab and have not moved an inch in 50 years. Zeno directed me to perform a sensitivity analysis on the stellar camera systems, and I published a classified paper in SCIF channels back in 1972. My guess is if it’s still extant, it’s still classified.

Notice the computer on the left side is an “SDS” mini-computer. That was built by Scientific Data Systems that later changed the name of the company to Digital Equipment Corporation (DEC). The reason was that “SDS” was also the abbreviation for “Student for a Democratic Society,” an infamous radical left-wing group responsible for numerous violent conflicts with law enforcement back in the early 1970s. DEC was late purchased by Dell Computer Corp.

The operator in the picture is Mr. Scott Rae, a staff Mathematician that also worked for Zeno.
In 1944, D.R. Hendrikz of the South African Trigonometrical Survey wrote, “For the computation of the geographical coordinates of the stations of the Geodetic Survey, Sir David Gill adopted the numerical values of the semi-major and semi-minor axes of Clarke’s 1880 figure or $a = 20,926,202$ ft and $b = 20,854,895$ ft. At that time this result was the most recent determination of the figure of the Earth. But, because the baselines were reduced to S.A.G. (South African Geodetic – ed.) feet, the computations were really carried out on a ‘Modified Clarke 1880 Spheroid’ defined by $a = 6,378,249.145$ 326 int. metre and $b = 6,356,514.966$ 721 int. metre. It may be remarked, in passing, that this value of the flattening for this spheroid is $1/f = 293.466$ 307 656, which differs slightly from the value 293.465 given by Clarke himself.” Later in the document, Hendrikz went on to present relations of the “Geodetic Cape rood” = 12.396 S.A.G. feet, and 1 Cape morgen = 600 square Cape roods = 2.116 539 816 acres. Note that the acre was originally the amount of English land that could be plowed in one day, and the morgen was roughly the amount of German land that could be plowed in a morning. Hendrikz stated that 1 Rhynland morgen = 0.634 282 acres.

Thanks to Professor Charles L. Merry of the University of Cape Town, “The ellipsoid is the Clarke 1880, oriented using astronomic observations of latitude, longitude and azimuth at a point near Port Elizabeth in the late 1800s. The offset from the geocentre is about 350m. An unusual feature is that it is the so-called ‘modified’ Clarke 1880 ellipsoid, because the conversion factor ‘yard-to-legal metre’ was used to convert Clarke’s values to international metres. The legal metre is based upon a defined relationship between the toise and the metre, not a physical standard, and is about 13ppm larger. Nevertheless, the official length standard is the international metre.

“Although local grid systems were common in the 19th and early 20th century, since the 1920s the Transverse Mercator (Gauss-Krüger) system has been exclusively used. It uses 2-degree wide panels, scale factor of unity on the central meridian and no false origin. The coordinate axes are directed South and West (no northings and eastings for us southerners!), and are labeled $x$ and $y$ respectively. It is a legal requirement for all cadastral surveys to use this grid system, and the large and medium scale national map series also uses it. The military use the UTM system, overprinted on the standard map sheets (false northings as well as false easting). …Contrary to what DMA (now NIMA-ed.) believe, they do not use the Arc datum. The Arc datum is used in parts of East Africa. It is based upon the same initial point near Port Elizabeth and the same ellipsoid (modified Clarke 1880), but uses a single chain of triangulation extracted from the national networks of South Africa, … and Zimbabwe, more or less along the 30th meridian. Close to Port Elizabeth, it is practically identical to the Cape datum, but diverges as one moves away. A GPS network is in place in Zimbabwe and the control networks are being re-adjusted. Although no final decision has been taken, it is likely that Zimbabwe will also convert to the WGS84 around the same time that South Africa does.” (Personal communication, July 1997).
Zimbabwe Update

“Recently, Zimbabwe recognized CORS technology as an integral component for the prompt acquisition of spatial data. Consequently, the country engaged the EU and UNDP to fund the process of establishing CORS in the country. Initially, five GNSS CORS will be established primarily focusing on expediting the process of boundary mapping of farms to support security of tenure thus ensuring food security in line with the Sustainable Development Goals. Having realized the urgent need for establishing CORS in Zimbabwe, the readiness of the stakeholders and institutions which use geospatial data is yet to be evaluated. Although the primary rationale for establishing CORS in Zimbabwe was motivated by the need to regularize the land reform program to ensure tenure security, it is apparent that other stakeholders besides cadastral surveyors will use this technology.” Establishment of Continuously Operating Reference Stations (CORS) in Zimbabwe, Mlambo, R., Freeman, A., African Journal on Land Policy and Geospatial Sciences ISSN:2657-2664, Vol.3 No.3 (September 2020) 42.

In Memoriam

George Erio

1945-2021

George Erio passed away on August 8, 2021, after a battle with liver cancer. George was a lifelong supporter of ASPRS since the early 1970s where is served on the old Pacific Region Board as President.

George received his B.Sc. in Civil Engineering from the University of Illinois at Urbana-Champaign and worked as a surveyor in the summers. George then worked for California Department of Transportation (CALTRANS) as a Civil Engineer. George then took a year off work at CALTRANS and studied analytical photogrammetry under James Anderson and Frank Moffitt and received his M.Sc. in Photogrammetry in 1972. George’s research was initially in camera calibration with Professor James Anderson. Later he put his knowledge of analytical photogrammetry into practice in developing an analytic bundle adjustment program which later became a commercial product and used by the federal government and private companies. After leaving CALTRANS George started Eriotech, LLC which provided aerotrigulation services and consulting.

George worked hard and played hard. When George wasn’t working, he enjoyed his many hobbies including backpacking, camping, canoeing, target shooting, fishing, gardening and playing the guitar. He retired in Key Largo, Florida but remained interested in the Society and the profession.

George is survived by his wife, Vanneza Alverez Erio, his children Amy and William; brothers, Peter, Michael and Tom; sister, Mary; grandchildren, Isabella and Luke; and niece, Diana. He will be missed by his family, friends and colleagues.

George Erio’s obituary can be viewed at www.butlerfuneralhomes.com/obituaries/George-William-Erio?obId=22010207#obituaryInfo.
NEW ASPRS MEMBERS

ASPRS would like to welcome the following new members!

Clement Akumu, PhD
Stuart Babin
William Barna
Jessica Bobeck
Theodore R. Brandt
Justin Cave
Kevin Chappell
Lori D. Collins, PhD
Peter David
Adam Dolberry, CP
Mitchell Ayemere Eboigbe
David Hughes
Ben Janevic
Babu Madukkathanam, PhD
Amber Middleton
Ilham Yasmeen Mohd Ishak
Charles Mondello
David Nilosek, PhD
Zachary Nixon
Alexander Nunez
Alex Richard Olsen-Mikutowicz
Leonard Pool
Abishek Poudel
Jennifer Pramuk
Hari Rajan
Nayely Rolon-Gomez
John Saunders
Stephen Smith
Jerry Tagesstad

FOR MORE INFORMATION ON ASPRS MEMBERSHIP, VISIT HTTP://WWW.ASPRS.ORG/JOIN-NOW
<table>
<thead>
<tr>
<th>SUSTAINING MEMBERS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACI USA Inc.</td>
</tr>
<tr>
<td>Weston, Florida</td>
</tr>
<tr>
<td><a href="https://acicorporation.com/">https://acicorporation.com/</a></td>
</tr>
<tr>
<td>Member Since: 2/2018</td>
</tr>
<tr>
<td>Aerial Services, Inc.</td>
</tr>
<tr>
<td>Cedar Falls, Iowa</td>
</tr>
<tr>
<td><a href="http://www.AerialServicesInc.com">www.AerialServicesInc.com</a></td>
</tr>
<tr>
<td>Member Since: 5/2001</td>
</tr>
<tr>
<td>Ayres Associates</td>
</tr>
<tr>
<td>Madison, Wisconsin</td>
</tr>
<tr>
<td><a href="http://www.AyresAssociates.com">www.AyresAssociates.com</a></td>
</tr>
<tr>
<td>Member Since: 1/1953</td>
</tr>
<tr>
<td>Dewberry</td>
</tr>
<tr>
<td>Fairfax, Virginia</td>
</tr>
<tr>
<td><a href="http://www.dewberry.com">www.dewberry.com</a></td>
</tr>
<tr>
<td>Member Since: 1/1985</td>
</tr>
<tr>
<td>Environmental Research Incorporated</td>
</tr>
<tr>
<td>Linden, Virginia</td>
</tr>
<tr>
<td><a href="http://www.eri.us.com">www.eri.us.com</a></td>
</tr>
<tr>
<td>Member Since: 8/2008</td>
</tr>
<tr>
<td>Esri</td>
</tr>
<tr>
<td>Redlands, California</td>
</tr>
<tr>
<td><a href="http://www.esri.com">www.esri.com</a></td>
</tr>
<tr>
<td>Member Since: 1/1987</td>
</tr>
<tr>
<td>GeoCue Group</td>
</tr>
<tr>
<td>Madison, Alabama</td>
</tr>
<tr>
<td><a href="http://www.geocue.com">http://www.geocue.com</a></td>
</tr>
<tr>
<td>Member Since: 10/2003</td>
</tr>
<tr>
<td>Geographic Imperatives LLC</td>
</tr>
<tr>
<td>Centennial, Colorado</td>
</tr>
<tr>
<td><a href="http://geographicimperativesllc.com">http://geographicimperativesllc.com</a></td>
</tr>
<tr>
<td>Member Since: 9/2021</td>
</tr>
<tr>
<td>GeoWing Mapping, Inc.</td>
</tr>
<tr>
<td>Richmond, California</td>
</tr>
<tr>
<td><a href="http://www.geowingmapping.com">www.geowingmapping.com</a></td>
</tr>
<tr>
<td>Member Since: 12/2016</td>
</tr>
<tr>
<td>GPI Geospatial Inc.</td>
</tr>
<tr>
<td>formerly Aerial Cartographics of America, Inc. (ACA)</td>
</tr>
<tr>
<td>Orlando, Florida</td>
</tr>
<tr>
<td><a href="http://www.aca-net.com">www.aca-net.com</a></td>
</tr>
<tr>
<td>Member Since: 10/1994</td>
</tr>
<tr>
<td>Green Grid Inc.</td>
</tr>
<tr>
<td>San Ramon, California</td>
</tr>
<tr>
<td><a href="http://www.greengridinc.com">www.greengridinc.com</a></td>
</tr>
<tr>
<td>Member Since: 1/2020</td>
</tr>
<tr>
<td>Halfff Associates, Inc.</td>
</tr>
<tr>
<td>Richardson, Texas</td>
</tr>
<tr>
<td><a href="http://www.halfff.com">www.halfff.com</a></td>
</tr>
<tr>
<td>Member Since: 8/2021</td>
</tr>
<tr>
<td>Keystone Aerial Surveys, Inc.</td>
</tr>
<tr>
<td>Philadelphia, Pennsylvania</td>
</tr>
<tr>
<td><a href="http://www.kasurveys.com">www.kasurveys.com</a></td>
</tr>
<tr>
<td>Member Since: 1/1985</td>
</tr>
<tr>
<td>Kucera International</td>
</tr>
<tr>
<td>Willoughby, Ohio</td>
</tr>
<tr>
<td><a href="http://www.kucerainternational.com">www.kucerainternational.com</a></td>
</tr>
<tr>
<td>Member Since: 1/1992</td>
</tr>
<tr>
<td>L3Harris Corporation</td>
</tr>
<tr>
<td>Broomfield, Colorado</td>
</tr>
<tr>
<td><a href="http://www.harris.com">www.harris.com</a></td>
</tr>
<tr>
<td>Member Since: 6/2008</td>
</tr>
<tr>
<td>NV5 Geospatial</td>
</tr>
<tr>
<td>Sheboygan Falls, Wisconsin</td>
</tr>
<tr>
<td><a href="http://www.quantumspatial.com">www.quantumspatial.com</a></td>
</tr>
<tr>
<td>Member Since: 1/1974</td>
</tr>
<tr>
<td>Pickett and Associates, Inc.</td>
</tr>
<tr>
<td>Bartow, Florida</td>
</tr>
<tr>
<td><a href="http://www.pickettusa.com">www.pickettusa.com</a></td>
</tr>
<tr>
<td>Member Since: 4/2007</td>
</tr>
<tr>
<td>Riegl USA, Inc.</td>
</tr>
<tr>
<td>Orlando, Florida</td>
</tr>
<tr>
<td><a href="http://www.rieglusa.com">www.rieglusa.com</a></td>
</tr>
<tr>
<td>Member Since: 11/2004</td>
</tr>
<tr>
<td>Robinson Aerial Surveys, Inc.(RAS)</td>
</tr>
<tr>
<td>Hackettstown, New Jersey</td>
</tr>
<tr>
<td><a href="http://www.robinsonaerial.com">www.robinsonaerial.com</a></td>
</tr>
<tr>
<td>Member Since: 1/1954</td>
</tr>
<tr>
<td>Sanborn Map Company</td>
</tr>
<tr>
<td>Colorado Springs, Colorado</td>
</tr>
<tr>
<td><a href="http://www.sanborn.com">www.sanborn.com</a></td>
</tr>
<tr>
<td>Member Since: 10/1984</td>
</tr>
<tr>
<td>Scorpius Imagery Inc.</td>
</tr>
<tr>
<td>Newark, Delaware</td>
</tr>
<tr>
<td><a href="mailto:aerial@scorpiusimagery.com">aerial@scorpiusimagery.com</a></td>
</tr>
<tr>
<td>Member Since: 6/2021</td>
</tr>
<tr>
<td>Surveying And Mapping, LLC (SAM)</td>
</tr>
<tr>
<td>Austin, Texas</td>
</tr>
<tr>
<td><a href="http://www.sam.biz">www.sam.biz</a></td>
</tr>
<tr>
<td>Member Since: 12/2011</td>
</tr>
<tr>
<td>T3 Global Strategies, Inc.</td>
</tr>
<tr>
<td>Bridgeville, Pennsylvania</td>
</tr>
<tr>
<td><a href="https://t3gs.com/">https://t3gs.com/</a></td>
</tr>
<tr>
<td>Member Since: 6/2020</td>
</tr>
<tr>
<td>Terra Remote Sensing (USA) Inc.</td>
</tr>
<tr>
<td>Bellevue, Washington</td>
</tr>
<tr>
<td><a href="http://www.teraremote.com">www.teraremote.com</a></td>
</tr>
<tr>
<td>Member Since: 11/2016</td>
</tr>
<tr>
<td>Towill, Inc.</td>
</tr>
<tr>
<td>San Francisco, California</td>
</tr>
<tr>
<td><a href="http://www.towill.com">www.towill.com</a></td>
</tr>
<tr>
<td>Member Since: 1/1952</td>
</tr>
<tr>
<td>Woolpert LLP</td>
</tr>
<tr>
<td>Dayton, Ohio</td>
</tr>
<tr>
<td><a href="http://www.woolpert.com">www.woolpert.com</a></td>
</tr>
<tr>
<td>Member Since: 1/1985</td>
</tr>
</tbody>
</table>

**SUSTAINING MEMBER BENEFITS**

**Membership**

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

**Benefits of an ASPRS Membership**

- Complimentary and discounted Employee Membership*
- E-mail blast to full ASPRS membership*
- Professional Certification Application fee discount for any employee
- 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
- Discount on PE&RS advertising
- 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
System Calibration Including Time Delay Estimation for GNSS/INS-Assisted Pushbroom Scanners Onboard UAV Platforms

Lisa M. LaForest, Tian Zhou, Seyyed Meghdad Hasheminasab, and Ayman Habib

Abstract
Unmanned aerial vehicles (UAVs) equipped with imaging sensors and integrated global navigation satellite system/inertial navigation system (GNSS/INS) units are used for numerous applications. Deriving reliable 3D coordinates from such UAVs is contingent on accurate geometric calibration, which encompasses the estimation of mounting parameters and synchronization errors. Through a rigorous impact analysis of such systematic errors, this article proposes a direct approach for spatial and temporal calibration (estimating system parameters through a bundle adjustment procedure) of a GNSS/INS-assisted pushbroom scanner onboard a UAV platform. The calibration results show that the horizontal and vertical accuracies are within the ground sampling distance of the sensor. Unlike for frame camera systems, this article also shows that the indirect approach is not a feasible solution for pushbroom scanners due to their limited ability for decoupling system parameters. This finding provides further support that the direct approach is recommended for spatial and temporal calibration of UAV pushbroom scanner systems.

Introduction
Remote sensing is a highly effective technique for gathering information without coming into contact with the object being mapped and is used for a variety of applications. One major application is in digital agriculture. Specifically, remote sensing systems onboard unmanned aerial vehicles (UAVs) are used for high-throughput phenotyping, plant localization, and biomass prediction. Deriving accurate 3D coordinates from such UAVs is contingent on accurate geometric calibration, which encompasses the estimation of mounting parameters and synchronization errors (Herwitz et al. 2004; Chen et al. 2017; Masjedi et al. 2018; Ahmed et al. 2019; Xu et al. 2019; Hasheminasab et al. 2020; Maimaitijiang et al., 2020). These systems have been proven as effective tools that are more accurate and cost effective than manual field measurements. Most UAV systems use RGB frame camera imagery. However, because of its wide range of spectral bands, hyperspectral pushbroom scanner imagery is also used for phenotyping and biomass prediction as well as soil content estimation and mineralogy (Haboudane et al. 2004; McBratney et al. 2005; Mulla 2013; Zhao et al. 2018; Notesco et al. 2020). For some other applications, such as classifying complex forest areas, modeling vascular plant species richness, and mapping coastal environments, LiDAR sensors have been adopted together with hyperspectral scanners to improve the synergy within the collected data (Dalponte et al. 2008; Sankey et al. 2017; Hakkenberg et al. 2018; Lin et al. 2019). For any remote sensing application, geometric system calibration is vital for determining accurate positions.

Geometric calibration deals with the estimation of any system parameters that might impact the positional accuracy of derived products. Therefore, it encompasses both intrinsic/extrinsic sensor parameters and synchronization error estimation (in this article referred to as spatial and temporal calibration, respectively). Spatial calibration accurately estimates both internal sensor characteristics and mounting parameters of the system components. Previous work has investigated spatial calibration of frame and hyperspectral pushbroom scanner systems (Lenz et al. 2014; Zhang et al. 2015; Wendel and Underwood 2017; Habib et al. 2018; Zhou et al. 2020). Another important aspect of geometric calibration of remote sensing systems is temporal calibration, which is the focus of this study. In order to ensure accurate derivation of platform position and orientation at the time of image exposure, these epochs have to be precisely time stamped with the corresponding global navigation satellite system (GNSS) time. Temporal calibration eliminates the time delay (i.e., synchronization errors) between the recorded timestamp by the integrated GNSS/inertial navigation system (GNSS/INS) unit of the image exposure epoch (usually referred to as event marker) and the actual time of image acquisition. A few groups that have focused on temporal calibration in the application of simultaneous localization and mapping. Li and Mourikis (2013) use an extended Kalman filter–based strategy by including time delay in the extended Kalman filter state vector. Furgale et al. (2013) adopt a least squares approach to estimate the spatial displacement and temporal offset between different sensors. Similar work has been conducted by Vo et al. (2016) and Nguyen and Lhuillier (2017). With regard to UAV-based systems, Elbahnasawy (2018) and Habib et al. (2018) focus on two hardware solutions to perform temporal calibration. The authors introduce a simulated feedback approach where a signal is simultaneously sent to both the GNSS/INS unit and the imaging system calibration.
sensor. However, this approach does not account for the camera response delay (i.e., the time delay between the camera receiving the signal and the camera capturing the image). The authors’ second approach, direct feedback, utilizes the camera flash hot shoe to generate a signal at the time the image is captured. This camera feedback signal is then sent to the onboard GNSS/INS unit, and a corresponding event marker is recorded, thus accounting for the camera response delay between receiving the triggering signal and actual image exposure.

While some researchers focus on hardware solutions for temporal calibration, others consider software-based approaches. Both Gabrlik et al. (2018) and Chiang et al. (2015) use a two-step software-based approach for estimating and removing time delay. First, exterior orientation parameters are estimated using indirect georeferencing with the help of ground control points (GCPs). Next, the difference between estimated exterior orientation parameters and GNSS/INS position and orientation information at the time of exposure is evaluated and used to solve for the lever arm components, boresight angles, and time delay. Finally, the calibration parameters, together with the GNSS/INS position and orientation information, are used for direct georeferencing of the sensors onboard the platform. Both Gabrlik et al. (2018) and Chiang et al. (2015) depend on GCPs and a two-step approach to estimate time delay in their systems. Blazquez (2008) presents a one-step approach for temporal calibration while using local geodetic coordinate systems. This approach focuses on adjusting the sensor model to include time delay as an additional system calibration parameter while using the GNSS/INS linear/angular velocity to consider the impact of time delay on the position and orientation information. This approach also depends on the availability of GCPs for accurate time delay estimation. Based on the one-step approach proposed by Blazquez (2008), Blazquez and Colomina (2012) expand the calibration model to handle global, compound mapping-geodetic coordinate systems by considering only GNSS/INS linear velocity. Rehak and Skaloud (2017) have also addressed temporal calibration. First, they use the direct feedback hardware approach to reduce time delay in their system to a minimum. Then they test two separate methods for temporal calibration. For the first method, they conduct a two-step approach that analyzes the residuals between the observed camera position and orientation by the GNSS/INS unit and those estimated through indirect georeferencing. Their second strategy, which requires GCPs, is based on a one-step approach that modifies the mathematical model to handle global, compound mapping-geodetic coordinate systems by considering only GNSS/INS linear velocity. An optimal flight configuration is then heuristically argued for. Furthermore, their study recommends that the lever arm components are estimated in a laboratory setting rather than during an in situ system calibration. LaForest et al. (2019) present two separate one-step approaches. The first one, denoted as the direct approach, modifies the existing mathematical model to include time delay as an additional system parameter. An optimal flight configuration is rigorously derived to suggest the most ideal collection geometry for simultaneously estimating spatial and temporal system calibration parameters. The second strategy, denoted as the indirect approach, exploits the optimal flight configuration and potential system parameter correlations to evaluate the time delay indirectly from the platform speed and estimated lever arm component in the flying direction. Both approaches successfully estimate time delay and mounting parameters without using GCPs.

Compared to previous work that focused only on frame camera systems, this article introduces and performs spatial and temporal calibration for pushbroom scanner systems. The key motivation for considering time delay in the geometric calibration of pushbroom scanners is the increase in the use onboard UAVs for high-throughput phenotyping. Due to the large-scale/fine-detail requirement of phenotyping applications and the fact that the integration strategies of such scanners with an onboard GNSS/INS unit cannot eliminate potential synchronization errors, we need to have access to precise yet practical calibration strategies. To address such a need, this article aims at highlighting the imposed challenges by pushbroom scanners when compared to frame camera systems during the spatial and temporal calibration process. Examples of such challenges include the impact of (1) 1D exposure of pushbroom scanners when compared with 2D coverage of frame cameras, (2) narrow angular field of view (AFOV) of used pushbroom scanners onboard UAVs, and (3) conflicting requirements of correlation maximization of the time delay/lever arm component along the flight direction and optimal configuration for the indirect approach. The article proceeds with an introduction of the methodology for geometric system calibration, including time delay estimation, using direct and indirect approaches. A discussion of an optimal flight configuration for the proposed direct approach while highlighting the incurred differences by the imaging mechanism of frame camera and pushbroom scanner systems will be also presented. Next, a description of the imaging platforms and collected data sets for the experimental analysis is presented. The results and analysis are then introduced for each system and experiment. The article concludes with a summary of its findings and directions for future research.

Methodology for System Calibration Including Time Delay Estimation and Optimal Flight Configuration

This section begins with a review and discussion of the collinearity equations and bundle block adjustment procedure for both frame and pushbroom scanner systems. Next, the direct approach using the modified mathematical model will be covered for pushbroom scanner systems. A detailed discussion of the optimal flight configuration for the direct system calibration approach for both frame cameras and pushbroom scanners will follow along with a comparison of the two. Finally, a discussion of the potential application of the indirect approach for pushbroom scanner systems concludes this section.

Collinearity Equations for Frame Camera and Pushbroom Scanner Systems

For remote sensing imaging platforms, photogrammetric techniques are used to derive 3D coordinates from acquired imagery. The collinearity equations, which represent the relationship between the 2D image coordinates and corresponding 3D ground coordinates, constitute the commonly used mathematical model. To express these equations, the following notations are used throughout this article: a vector connecting point $b$ to point $a$ relative to a coordinate system associated with point $b$ is represented as $r_b^a$, and a rotation matrix transforming a vector from coordinate system $a$ to another coordinate system $b$ is represented as $R_{b}^{a}$. Using such notations, the collinearity equations expressing the 3D ground coordinates as a function of the corresponding 2D image coordinates are shown in Equation 1 (Schwarz et al. 1993):

$$r_i^m = r_{i(0)}^m + R_{i(0)}^m \rho_b^a + \lambda (i,t) R_{i(0)}^m r_i^{(0)}$$

where $r_i^m$ are ground coordinates of object point $i$, $r_{i(0)}^{(0)}$ is the coordinate of image point $i$ captured by the camera at time $t$ relative to the camera coordinate system, $t$ is the actual time of exposure, $R_{i(0)}^m$ is the position of the inertial measurement unit (IMU) body frame relative to the mapping reference frame at time $t$ derived through the GNSS/INS integration process, $R_{i(0)}^m$ is the rotation matrix relating the IMU body frame and mapping reference frame at time $t$ derived through the GNSS/INS integration process, $r_b^a$ is the lever arm describing the position of the camera relative to the IMU body frame, $R_{b}^{a}$ is the rotation matrix relating the
camera coordinate system and IMU body frame, and \( \lambda(i, t) \) is the scale factor for image point \( i \) captured by the camera at time \( t \).

Equation 1 is a generic form of the collinearity equations and can be used for frame camera and pushbroom scanner imaging systems. Henceforth, a scene is considered as a 2D coverage area of the ground. The main difference between frame cameras and pushbroom scanners is that the frame camera’s scene is captured in one instance by a single image, whereas the pushbroom scanner’s scene is built by multiple images captured sequentially. Given this geometry, the \( x \) and \( y \) values in the image coordinate vector, \( r_{ib}^{(0)} \), for a frame camera system will be variable, depending on the image point location, which is bounded by the sensor’s AFOV. However, pushbroom scanner’s \( x \) image coordinates will have variable values, but \( y \) will always be constant, depending on the pushbroom scanner alignment on the sensor’s focal plane (note that the \( x \) axis of the image coordinate system is defined along the scan line).

### Direct Approach for Time Delay Estimation in Pushbroom Scanner Systems

The direct approach for GNSS/INS-assisted frame camera systems was first presented by Blazquez (2008), where the mathematical model was modified to incorporate time delay as an additional system parameter. This section introduces the modified mathematical model for frame cameras and pushbroom scanner systems. Examining Equation 1, there are three terms that are affected by a time delay: position, \( r_{ib}^{(0)} \), of the platform; orientation, \( R_{bi}^{(0)} \), of the platform; and scale, \( \lambda(i, t) \). However, the scale factor is routinely eliminated by dividing the first and second equations by the third after expressing the image coordinates as a function of the ground coordinates, leaving only the position and orientation as the components directly affected by the time delay. The goal is to modify the mathematical model in Equation 1 to account for potential time delay affecting the position and orientation components. First, definitions of the different times involved are provided. Throughout this article, we will refer to the term “event marker,” which is used to indicate the time of exposure based on the feedback signals received by the GNSS/INS unit from the camera. The initial GNSS/INS event marker time is denoted by \( t_0 \). The time delay is denoted by \( \Delta t \). Finally, combining those terms, the actual sensor exposure time is expressed as \( t \), which is equal to \( t_0 + \Delta t \). Using those time definitions, the position that accounts for time delay, \( r_{ib}^{(0)} \), can be expressed by the position at the initial event marker, \( r_{ib}^{(0)} \), and adding the displacement caused by the time delay, \( \Delta r_{ib} \), where \( \Delta r_{ib} = \dot{r}_{ib} \Delta t \), is the instantaneous linear velocity at \( t_0 \), and is comprised of three components, \( \dot{r}_{ib}^{(0)}(x) \), \( \dot{r}_{ib}^{(0)}(y) \), and \( \dot{r}_{ib}^{(0)}(z) \), in the \( x \), \( y \), and \( z \) directions, respectively. The second term affected by a time delay is the GNSS/INS-based orientation of the IMU body frame, \( R_{bi}^{(0)} \). The change in the IMU orientation caused by the time delay, \( R_{bi}^{(0)} \), is expressed by the incremental rotation matrix in Equation 2, where \( \omega_{bi}^{(0)} \), \( \phi_{bi}^{(0)} \), and \( \kappa_{bi}^{(0)} \), are the instantaneous pitch, roll, and heading angular velocity at \( t_0 \), respectively. As a result, the orientation of the IMU body frame at the actual exposure time, \( R_{bi}^{(0)} \), is represented by the multiplication of the orientation at the initial event marker, \( R_{bi}^{(0)} \), with the attitude change caused by the time delay, \( R_{bi}^{(0)} \). The instantaneous linear/angular velocity is estimated through evaluating the change in position/orientation parameters within a user-specified time interval. More details can be found in LaForest et al. (2019):

\[
R_{bi}^{(0)}(t_0 + \Delta t) = \text{Rotation}(\omega_{bi}^{(0)} \Delta t, \phi_{bi}^{(0)} \Delta t, \kappa_{bi}^{(0)} \Delta t)
\]

The collinearity equations expressed in Equation 1 can now be modified to include the time delay as a system parameter and are expressed in Equation 3.

\[
r_i^* = \frac{1}{\lambda(i, t)} R_i^b \left[ R_{bi}^{(0)} R_{bi}^{(0)} \left[ r_i^{(b_0)} - r_{bi}^{(0)} - e_{bi}^{(0)} \Delta t \right] - r_{bi}^{(0)} \right] \tag{3}
\]

### Optimal Flight Configuration for Pushbroom Scanner System Calibration While Considering Time Delay

A control-free system calibration requires an optimal flight configuration that maximizes the impact of deviations in the system parameters from the true ones (such deviations will be denoted as systematic errors or biases). An optimal flight configuration allows for the decoupling of the systematic errors; hence, they can be more easily detected and quantified. In other words, an optimal configuration would provide the minimal configuration that allows for accurate estimation of the system calibration parameters. To establish the optimal flight configuration, a bias impact analysis of system parameters on the derived ground coordinates is carried out. The bias impact analysis is conducted by examining the partial derivatives of the ground coordinates in the collinearity equations with respect to each system parameter. A few assumptions are made for deriving the partial derivatives. These assumptions are intended only to simplify the derivations. The assumptions are that (1) the sensor and IMU body frame coordinate systems are almost parallel, leading to small boresight angles; (2) the sensor and IMU body frame coordinate systems are vertical; (3) the sensor is traveling with a constant attitude in the south-to-north and north-to-south directions; (4) the terrain is horizontal and flat; and (5) the lever arm components are relatively small. For situations where the sensor and IMU body frame coordinate systems are not almost parallel, one can introduce a virtual sensor coordinate system that is parallel to the IMU body frame. Deviations from the remaining assumptions would have a more favorable effect on the ability to decouple the various system parameters. In other words, such assumptions represent the worst-case scenario that could negatively impact the separability of the system calibration parameters.

To evaluate the partial derivatives, the modified collinearity model expressed in Equation 3 is reformulated to represent the ground coordinates as a function of the system parameters and image measurements, shown in Equation 4. In this equation, the three terms, comprised of system parameters, needed for the bias impact analysis are \( \Delta R_i^b, R_{bi}^{(0)^T}, \) and \( \lambda(i, t) \).
The partial derivatives are then derived using the assumptions made above and are expressed in Table 1 for both frame camera and pushbroom scanner systems. Several terms in Table 1 are preceded by a double sign (± or τ). In such a case, the top and bottom signs refer to the south-to-north and north-to-south flight lines, respectively. As mentioned earlier, the main difference between frame cameras and pushbroom scanners is that for the latter, the yi image measurements are equal to zero (assuming that the pushbroom scanner is mounted vertically below the perspective center). This difference affects several system parameters’ impact on ground coordinates, introducing more challenges in decoupling the different terms:

\[
\begin{align*}
\hat{r}_i^{m} &= \hat{r}_i^{m}(t)R_{\beta(i)}^mR_{\alpha(i)}^m + R_{\beta(i)}^m R_{\alpha(i)}^m \hat{r}_i^{m} + \lambda (i,t) R_{\beta(i)}^m R_{\alpha(i)}^m R_{\gamma(i)}^m \hat{r}_i^{m} \\
&= \hat{r}_i^{m} + \Delta \hat{r}_i^{m}(t) + \Phi \hat{r}_i^{m}(t) + \Psi \hat{r}_i^{m}(t) + \Omega \hat{r}_i^{m}(t) + \Theta \hat{r}_i^{m}(t)
\end{align*}
\]  

(4)

The partial derivatives are now used to assess the impact of each system parameter on the 3D point coordinates for frame camera and pushbroom scanner systems. The focus will be on the system parameters that have the possibility of being highly correlated or more difficult to decouple given a specific flight configuration, imaging mechanism, and image point distribution. The conceptual basis of the bias impact analysis and derivation of optimal flight configuration is identifying the flight pattern and image point distribution that would impact the target function of the BA procedure (i.e., the one that would negatively impact the intersection quality of conjugate light rays from such a flight configuration in the presence of biases in the system parameters). Following are the findings that can be drawn from the analysis of the reported bias impacts in Table 1:

1. As can be seen in row 3 of Table 1, a \(\delta Z\) bias in the lever arm will have a constant impact (\(\delta Z\)) regardless of the flight configuration and image point location. In other words, the impact of a bias in the \(Z\) component of the lever arm will not affect the quality of multi-light ray intersection within the BA. Therefore, the lever arm component in the \(Z\) direction cannot be estimated unless vertical control is available (in such a case, the photogrammetric surface will show a bias, \(\delta Z\), when compared to the defined surface by the vertical control).

2. Examining rows 2 and 4 of Table 1, a difference between frame cameras and pushbroom scanners can be observed in the \(Y\) direction. For frame cameras, a \(\delta Y\) bias in the lever arm affects the \(Y\) direction only; on the other hand,

### Table 1. Impact of bias in system parameters on 3D point coordinates for frame camera and pushbroom scanner systems.

<table>
<thead>
<tr>
<th>Bias in System Parameters</th>
<th>Impact on Ground Coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frame</td>
</tr>
<tr>
<td>1. (\delta \Delta X)</td>
<td>(\pm \delta \Delta X)</td>
</tr>
<tr>
<td>2. (\delta \Delta Y)</td>
<td>(\pm \delta \Delta Y)</td>
</tr>
<tr>
<td>3. (\delta \Delta Z)</td>
<td>(\pm \delta \Delta Z)</td>
</tr>
<tr>
<td>4. (\delta \Delta \omega)</td>
<td>(\pm \frac{X_i Y_i}{c^2} H \delta \Delta \omega)</td>
</tr>
<tr>
<td>5. (\delta \Delta \varphi)</td>
<td>(\pm \frac{Y_i}{c} H \delta \Delta \varphi)</td>
</tr>
<tr>
<td>6. (\delta \Delta \kappa)</td>
<td>(\pm \frac{1}{c} H \delta \Delta \kappa)</td>
</tr>
</tbody>
</table>

### Table 1. continued

<table>
<thead>
<tr>
<th>Bias in System Parameters</th>
<th>Impact on Ground Coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frame</td>
</tr>
<tr>
<td>1. (\delta \Delta X)</td>
<td>0</td>
</tr>
<tr>
<td>2. (\delta \Delta Y)</td>
<td>(\pm \delta \Delta Y)</td>
</tr>
<tr>
<td>3. (\delta \Delta Z)</td>
<td>0</td>
</tr>
<tr>
<td>4. (\delta \Delta \omega)</td>
<td>(\pm \frac{1}{c} H \delta \Delta \omega)</td>
</tr>
<tr>
<td>5. (\delta \Delta \varphi)</td>
<td>(\pm \frac{X_i Y_i}{c^2} H \delta \Delta \varphi)</td>
</tr>
<tr>
<td>6. (\delta \Delta \kappa)</td>
<td>(\pm \frac{1}{c} H \delta \Delta \kappa)</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
\hat{r}_i^{m}(Y) \delta \Delta t + \lambda (i,t) \left[ \pm \left( 1 + \frac{Y_i^2}{c^2} \right) \phi_{\beta(i)} Y_i \pm \chi_i \right] \delta \Delta t \\
\hat{r}_i^{m}(Z) \delta \Delta t + \lambda (i,t) \left[ \pm \left( 1 + \frac{Z_i^2}{c^2} \right) \phi_{\gamma(i)} Z_i \pm \chi_i \right] \delta \Delta t
\end{align*}
\]
a \( \delta \omega \) bias in the boresight pitch angle affects both the \( X \) and the \( Y \) direction. Therefore, \( \delta Y \) and \( \delta \omega \) can be easily decoupled when dealing with frame camera imagery that have tie points with nonzero \( xy \) image coordinates. However, for pushbroom scanners, the \( \delta Y \) and \( \delta \omega \) biases in the lever arm and boresight pitch angle, respectively, affect only the \( Y \) direction. Moreover, the impact of both biases is constant for a given flying height and does not depend on the image point location. Such a characteristic of the bias impact leads to a high probability of \( \delta \omega \) being strongly correlated or harder to decouple unless flight lines with significant difference in the flight height are available. One should note that having UAV flights with significant elevation variations could be impractical.

3. Analyzing rows 1 and 5 of Table 1, for frame cameras, the \( \delta X \) bias in the lever arm affects the \( X \) direction only, while a \( \delta \phi \) bias in the boresight roll angle affects both the \( X \) and the \( Y \) direction. Due to such variability in the impact pattern, the \( \delta X \) and \( \delta \phi \) can be decoupled for frame cameras. However, for pushbroom scanners, the \( \delta X \) and \( \delta \phi \) biases in the lever arm and boresight roll angle, respectively, affect only the \( X \) direction. Therefore, this creates an increased probability of \( \delta X \) and \( \delta \phi \) being highly correlated or harder to decouple. This is especially true when there is a single flying height coupled with a small AFOV as defined by \( \chi/c \).

4. For situations where we have flight lines from a single (or similar) flying height(s) and/or pushbroom scanners with narrow AFOV, one should expect correlation among biases in the \( Y \) lever arm component, boresight pitch angle, and boresight heading angle (i.e., \( \delta \omega \)) being highly correlated or harder to decouple. In this regard, one should note that frame cameras will not have as much correlation among \( \delta Y \), \( \delta \omega \), and \( \delta \phi \), respectively. In this case, the biases \( \delta Y \) and \( \delta \phi \) affect both the \( X \) and \( Y \) direction. Moreover, the decoupling of \( \delta \omega \) and \( \delta \phi \) can rely on the inherently larger AFOV and variability in the image point coordinates.

5. Analyzing row 7 of Table 1, for frame camera and pushbroom scanner systems, it is important to have a variability in the linear/angular velocity components to ensure that the time delay can be precisely estimated. For frame camera systems, having variability in the linear velocity or any of the angular velocity components, \( \partial \) or \( \phi \), will ensure that the time delay is decoupled from other system parameters. However, for pushbroom scanner systems, since the \( y \) image coordinate is always 0, \( \partial \)'s impact on the \( X \) direction, that is, \( \partial Y \) is affected only by the \( \partial Y \) linear velocity component and the \( \phi \) angular velocity component. The \( \partial \)'s impact on the \( Y \) direction, that is, \( \partial Y \) is affected only by the \( \partial Y \) linear velocity component and \( \phi \) angular velocity components. Therefore, pushbroom scanner systems rely more on variability in the different components of angular velocity than do frame cameras. Moreover, if the pushbroom scanner system has a small AFOV, \( \chi \), variability is limited, further restricting the possibility of decoupling \( \delta Y \) and \( \delta \).

Given the above bias impact analysis, it can be concluded that the optimal flight configuration for pushbroom scanner systems is comprised of opposite flying directions at different flying heights. Moreover, variations in linear/angular velocity are recommended. Finally, good image point distribution along the scan line will help in decoupling the system parameters.

Potential Implementation of the Indirect Approach for Time Delay Estimation in Pushbroom Scanners

The bias impact analysis will now be investigated to check whether the indirect approach, which has been previously introduced for frame cameras, can be used for pushbroom scanners. For frame camera systems, the time delay is correlated with the lever arm component along the flying direction, \( \delta Y \), given a single ground speed and minor changes in linear/angular velocity, shown in Table 1 (rows 2 and 7). In the indirect approach, the time delay is not directly derived. More specifically, it is indirectly estimated using the lever arm deviation from the nominal value in the along-flying direction and speed/time/distance relation. Thus, for frame cameras, the indirect approach is meant for special cases in which an existing BA cannot be modified to incorporate time delay as an additional system parameter (LaForest et al. 2019).

To induce a correlation between the time delay and lever arm component along the flying direction, single flying speed and constant attitude are needed. At this stage, one should note that multiple flying heights are necessary when calibrating pushbroom scanners (even when excluding the time delay). The ground sampling distance (GSD) of the pushbroom scanner along/across flying direction depends on the flying height/speed, respectively. To ensure that the GSD in the along/across flying directions are equal, a specific flying speed is chosen at a given height. Thus, different flying heights would require different flying speeds. In summary, decoupling the mounting parameters and application of the indirect approach for pushbroom scanner system calibration have conflicting requirements. Decoupling the mounting parameters requires multiple flying heights to avoid: (1) a correlation between the lever arm component along the flight direction and boresight pitch angle, (2) a correlation between the lever arm component across the flight direction and boresight roll angle, and (3) introduced correlations by the small AFOV of most pushbroom scanners onboard UAV systems. Using the indirect approach, on the other hand, requires flights from a single flying height to maintain a single speed. Therefore, the indirect approach, when applied to pushbroom scanner scenes from a single flight height, would produce inaccurate estimates of the system parameters. The section on experimental results will show an example of applying the indirect approach for the calibration of a pushbroom scanner system to emphasize this conclusion.

System Specifications and Data Set Description

Data Acquisition Systems

In this study, several platforms equipped with a GNSS/INS unit and different imaging sensors have been used. The DJI Matrice 600 Pro (M600P) was the UAV platform used. This system has been equipped with the Applanix APX-15 UAV V2 GNSS/INS unit for direct georeferencing with a predicted positional accuracy of 2 to 5 cm and heading and roll/pitch accuracies of 0.08° and 0.25°, respectively (APX 2017). The DJI M600P platforms have been also equipped with three different pushbroom scanner sensors. Two of the pushbroom scanners used were the Headwall’s Nano-Hyperspec sensor operating at the visible and near-infrared bands. In this study, the Nano-Hyperspec sensors will be referred to as nHS, along with their corresponding ending serial numbers, that is, nHS-70 and nHS-199. These Nano-Hyperspec sensors cover 270 to 273 nm, corresponding to a detector pitch of 7.5 μm (Headwall Photonics, Inc. 2018b). The nHS cameras have a focal length of 8.2 mm, resulting in a field of view of approximately 31° (horizontal). The third pushbroom scanner aboard the DJI M600P was the shortwave infrared sensor, referred to as uVS-307 in this study. The uVS-307 covers 166 spectral bands ranging between 900 and 2500 nm with a bandwidth of 9.6 nm. The 1D array of the uVS-307 consists of 640 pixels and has a pixel pitch of 24 μm (Headwall Photonics, Inc. 2018a). The focal length of the uVS-307 is 24.6 mm, resulting in a field of view of 21° (horizontal). The nHS-70 and uVS-307 were installed on the same platform. For the nHS-199, the UAV was also equipped with both a LiDAR and an RGB imaging sensor. However, for this study, the
nHS-199 and APX unit were the only sensors used during data collection. An image of the DJI M600P equipped with the nHS-70 and uVS-307, along with the sensor and vehicle coordinate systems orientation, is shown in Figure 1a. An image of the DJI M600P equipped with the nHS-199 along with the sensor and vehicle coordinate systems alignment is shown in Figure 1b. Table 2 outlines the nominal lever arm and boresight values for both systems. It is worth mentioning that the X lever arm component is aligned along the flying direction (highlighted in bold in Table 2) for these systems. Hence, for the indirect approach, derived $\Delta X$ rather than $\Delta Y$ will be used to estimate the time delay. As mentioned earlier, all three pushbroom scanner systems onboard the DJI M600P are equipped with the APX. The system records a pulse per second, frame index count per line, position and orientation along with the coordinated universal time (UTC at a 100-Hz data rate), and Nano/shortwave infrared clock count. The link between the pulse per second from the APX, frame index count, and UTC time is used to establish the time stamp for each scan line.

### Data Set Description
In this study, a total of four data collection missions have been carried out using the introduced systems in the previous section. Table 3 shows the data acquisition date, sensor, and corresponding flight parameters for each collection. For all the collection missions, there are between four and five checkerboard targets that are used as tie features. The checkerboard targets are measured in the original hyperspectral scenes with the number of measured image points also listed in Table 3. Figure 2 shows an example of the configuration of checkerboard targets and flight lines for the 9 July 2019 nHS-70/uVS-307 data collection. The configuration of checkerboard targets and flight lines are similar for all other collection missions. The five checkerboard targets were surveyed by a Topcon GR-5 GNSS receiver with an accuracy of 2 to 3 cm. A sample of the raw uVS-307 and nHS-70 imagery over the calibration field is shown in Figure 3a and 3b. Figure 4a and 4b show the distribution of the tie point scene coordinates of the uVS-307 9 July and 14 August 2009 data sets, respectively. It can be observed that the 9 July data set has an unfavorable tie point distribution compared to the 14 August data set (i.e., the former has less variability in the scene coordinates along the scan line than the latter). As for the data sets from nHS-70 and...
nHS-199, the distribution of the tie point scene coordinates is similar to the uVS-307 14 August data set.

Experimental Results and Analysis

In this section, the experimental objectives and results are discussed. The first objective of the conducted experiments is to evaluate the direct approach’s ability to accurately estimate system calibration parameters and ground coordinates of checkpoints. The estimated system parameters include the X and Y components of the lever arm, boresight angles, and

Table 3. Flight parameters and number of measured image points for the different systems and collection dates.

<table>
<thead>
<tr>
<th>Sensor and Date</th>
<th>Flying Height (m)</th>
<th>Ground Speed (m/s)</th>
<th>Ground Sampling Distance (cm)</th>
<th>Sidelap (%)</th>
<th>No. of Flight Lines</th>
<th>No. of Measured Image Points (All/20 m/40 m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>nHS-70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 July 2019</td>
<td>20</td>
<td>3</td>
<td>1.79</td>
<td>60</td>
<td>8</td>
<td>64/30/34</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>3</td>
<td>3.58</td>
<td>80</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>14 August 2019</td>
<td>20</td>
<td>3</td>
<td>1.79</td>
<td>60</td>
<td>6</td>
<td>42/24/18</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>6</td>
<td>3.58</td>
<td>80</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>uVS-307</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 July 2019</td>
<td>20</td>
<td>3</td>
<td>1.95</td>
<td>42</td>
<td>8</td>
<td>28/10/18</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>6</td>
<td>3.89</td>
<td>71</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>14 August 2019</td>
<td>20</td>
<td>3</td>
<td>1.95</td>
<td>42</td>
<td>6</td>
<td>24/17/7</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>6</td>
<td>3.89</td>
<td>71</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>nHS-199</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29 May 2019</td>
<td>20</td>
<td>3</td>
<td>1.79</td>
<td>60</td>
<td>6</td>
<td>39/24/15</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>6</td>
<td>3.58</td>
<td>80</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>11 July 2019</td>
<td>20</td>
<td>3</td>
<td>1.79</td>
<td>60</td>
<td>10</td>
<td>46/24/22</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>6</td>
<td>3.58</td>
<td>80</td>
<td>8</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3. Sample of the raw (a) uVS-307 and (b) nHS-70 imagery over the calibration field with enhanced representation of the checkerboard targets.

Figure 4. Tie point scene coordinate distribution for the uVS-307 (a) 9 July and (b) 14 August data sets.
Table 4. Estimated system parameters and their standard deviation together with the square root of an a posteriori variance factor.

<table>
<thead>
<tr>
<th>Nominal Values</th>
<th>Δt (ms)</th>
<th>ΔX (m)</th>
<th>ΔY (m)</th>
<th>Δω (°)</th>
<th>Δφ (°)</th>
<th>Δκ (°)</th>
<th>Δf (Pixel)</th>
</tr>
</thead>
<tbody>
<tr>
<td>nHS-70</td>
<td>0</td>
<td>−0.02</td>
<td>0.06</td>
<td>0</td>
<td>0</td>
<td>180</td>
<td>N/A</td>
</tr>
<tr>
<td>uVS-307</td>
<td>0</td>
<td>−0.14</td>
<td>0.06</td>
<td>0</td>
<td>0</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>nHS-199</td>
<td>0</td>
<td>0.07</td>
<td>0.04</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Estimated System Parameters While Ignoring Time Delay

<table>
<thead>
<tr>
<th>nHS-70</th>
<th>9 July</th>
<th>N/A</th>
<th>0.011 ± 0.008</th>
<th>0.067 ± 0.008</th>
<th>0.211 ± 0.016</th>
<th>0.013 ± 0.017</th>
<th>179.059 ± 0.030</th>
<th>0.820</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>14 August</td>
<td>N/A</td>
<td>−0.002 ± 0.009</td>
<td>0.057 ± 0.009</td>
<td>0.175 ± 0.021</td>
<td>0.026 ± 0.023</td>
<td>−179.904 ± 0.041</td>
<td>0.897</td>
</tr>
<tr>
<td>uVS-307</td>
<td>9 July</td>
<td>N/A</td>
<td>−0.108 ± 0.010</td>
<td>0.060 ± 0.010</td>
<td>0.293 ± 0.020</td>
<td>0.419 ± 0.026</td>
<td>0.039 ± 0.701</td>
<td>0.604</td>
</tr>
<tr>
<td></td>
<td>14 August</td>
<td>N/A</td>
<td>−0.097 ± 0.013</td>
<td>0.040 ± 0.013</td>
<td>0.236 ± 0.034</td>
<td>0.371 ± 0.033</td>
<td>0.031 ± 0.109</td>
<td>0.777</td>
</tr>
<tr>
<td>nHS-199</td>
<td>29 May</td>
<td>N/A</td>
<td>0.062 ± 0.011</td>
<td>0.040 ± 0.011</td>
<td>0.289 ± 0.023</td>
<td>−1.323 ± 0.023</td>
<td>−0.450 ± 0.043</td>
<td>0.772</td>
</tr>
<tr>
<td></td>
<td>11 July</td>
<td>N/A</td>
<td>0.086 ± 0.013</td>
<td>0.035 ± 0.013</td>
<td>0.460 ± 0.029</td>
<td>−1.126 ± 0.030</td>
<td>−0.460 ± 0.064</td>
<td>1.245</td>
</tr>
</tbody>
</table>

Estimated System Parameters While Considering Time Delay (Direct Approach)

<table>
<thead>
<tr>
<th>nHS-70</th>
<th>9 July</th>
<th>5.912 ± 0.340</th>
<th>0.030 ± 0.006</th>
<th>0.066 ± 0.006</th>
<th>0.210 ± 0.012</th>
<th>−0.094 ± 0.015</th>
<th>179.977 ± 0.023</th>
<th>0.624</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>14 August</td>
<td>10.464 ± 0.851</td>
<td>0.007 ± 0.008</td>
<td>0.050 ± 0.008</td>
<td>0.163 ± 0.017</td>
<td>−0.098 ± 0.021</td>
<td>179.938 ± 0.035</td>
<td>0.727</td>
</tr>
<tr>
<td>uVS-307</td>
<td>9 July</td>
<td>4.146 ± 0.383</td>
<td>−0.097 ± 0.008</td>
<td>0.060 ± 0.007</td>
<td>0.295 ± 0.015</td>
<td>0.309 ± 0.028</td>
<td>0.031 ± 0.549</td>
<td>0.473</td>
</tr>
<tr>
<td></td>
<td>14 August</td>
<td>9.876 ± 0.903</td>
<td>−0.097 ± 0.009</td>
<td>0.037 ± 0.009</td>
<td>0.242 ± 0.022</td>
<td>0.273 ± 0.024</td>
<td>−0.021 ± 0.072</td>
<td>0.514</td>
</tr>
<tr>
<td>nHS-199</td>
<td>29 May</td>
<td>9.824 ± 0.788</td>
<td>0.079 ± 0.011</td>
<td>0.050 ± 0.011</td>
<td>0.316 ± 0.022</td>
<td>−1.351 ± 0.024</td>
<td>−0.480 ± 0.040</td>
<td>0.729</td>
</tr>
<tr>
<td></td>
<td>11 July</td>
<td>10.328 ± 0.530</td>
<td>0.109 ± 0.010</td>
<td>0.029 ± 0.009</td>
<td>0.451 ± 0.021</td>
<td>−1.315 ± 0.027</td>
<td>−0.471 ± 0.046</td>
<td>0.890</td>
</tr>
</tbody>
</table>

Table 5. Correlation matrix of system parameters for 9 July nHS-70 direct approach results (high correlation values, larger than ±0.85, are in bold).

<table>
<thead>
<tr>
<th></th>
<th>Δt</th>
<th>ΔX</th>
<th>ΔY</th>
<th>Δω</th>
<th>Δφ</th>
<th>Δκ</th>
<th>Δf</th>
</tr>
</thead>
<tbody>
<tr>
<td>Δt</td>
<td>1</td>
<td>−0.025</td>
<td>0.045</td>
<td>0.019</td>
<td>0.042</td>
<td>−0.080</td>
<td>−0.021</td>
</tr>
<tr>
<td>ΔX</td>
<td>−0.025</td>
<td>1</td>
<td>0.945</td>
<td>0.075</td>
<td>−0.041</td>
<td>−0.091</td>
<td>0.121</td>
</tr>
<tr>
<td>ΔY</td>
<td>0.945</td>
<td>0.075</td>
<td>1</td>
<td>−0.017</td>
<td>−0.091</td>
<td>0.480</td>
<td>0.271</td>
</tr>
<tr>
<td>Δω</td>
<td>0.019</td>
<td>0.042</td>
<td>−0.017</td>
<td>1</td>
<td>−0.071</td>
<td>−0.480</td>
<td>0.271</td>
</tr>
<tr>
<td>ΔΔφ</td>
<td>−0.080</td>
<td>0.042</td>
<td>−0.017</td>
<td>−0.071</td>
<td>1</td>
<td>−0.480</td>
<td>0.271</td>
</tr>
<tr>
<td>ΔΔκ</td>
<td>−0.021</td>
<td>−0.041</td>
<td>−0.017</td>
<td>−0.071</td>
<td>1</td>
<td>0.271</td>
<td>1</td>
</tr>
</tbody>
</table>

9 July data set. The correlations highlighted in bold are above a chosen ±0.85 threshold and are flagged as high. The ΔX lever arm component and Δφ boresight angle as well as the ΔY lever arm component and Δω boresight angle have the highest correlation. This is expected based on the previously discussed bias impact analysis. Furthermore, the time delay, Δt, is not correlated with any of the parameters, an indication of sufficient variability in the angular/linear velocity. The correlation matrices for the remaining data sets as well as scenarios ignoring time delay have similar results.

A summary of the mean standard deviation (relative accuracy measure) of the checkpoints derived from the BA as well as mean/standard deviation/root mean square error (RMSE) of the differences between the BA-based and surveyed coordinates for the checkpoints (absolute accuracy measure) for the nHS-70, uVS-307, and nHS-199 are shown in Table 6. The relative accuracy while considering/ignoring time delay for the horizontal components shows minimal differences. The vertical relative accuracy shows a slight improvement while considering time delay. The differences in horizontal components of the RMSE while considering/ignoring time delay do not exceed 2 cm. For the vertical component of the absolute accuracy measures, the difference in the RMSE while considering/ignoring time delay is 1.2 cm for the nHS-70 14 August data set (i.e., considering the time delay showed only a slight improvement in the vertical direction for this data set). However, the vertical component for the uVS-307 9 July data set shows a 28.1-cm improvement while considering/ignoring time delay. Given that the uVS-307 9 July data set

time delay. All imaging sensors—nHS-70, uVS-307, and nHS-199—and all the collection missions are included in testing this objective. The next objective is evaluating the ability of the direct approach to handle a larger time delay while accurately estimating system parameters and ground coordinates of checkpoints. To test this objective, an artificial time delay of 200 ms is introduced to the nHS-70 and nHS-199 data sets. A 200-ms time delay was chosen because previous work indicated that this was the amount found in other consumer-grade remote sensing platforms (LaForest et al. 2019). For the above objectives, a comparison while considering/ignoring time delay is conducted. For the different tests, the relative accuracy is evaluated by examining the derived standard deviation of checkpoint coordinates from the BA. The absolute accuracy is evaluated by comparing the checkpoint ground coordinates to the surveyed ones. The last objective, which is evaluated using the nHS-70 data set, tests the ability of the indirect approach to accurately estimate pushbroom scanner system parameters and ground coordinates of checkpoints.

Objective 1: Feasibility of Using the Direct Approach to Estimate Time Delay for Pushbroom Scanner Systems

The direct approach results for the nHS-70, uVS-307, and nHS-199 are presented below. The results include experiments while considering/ignoring time delay. The nominal values and estimated system parameters are shown in Table 4 along with the square root of the a posteriori variance factor σo, which gives an indication of the magnitude of back-projection errors for the different tie points. The time delay estimated for each system and collection mission ranges between 4.1 and 10.4 ms. A time delay of this magnitude is considered relatively small. The ΔX and ΔY lever arm components and boresight angles are also estimated during these calibration tests. As presented in Table 4, σo values for all the data sets are within 1 pixel, which reveals that the estimated system parameters produced a back-projection error in the same range of image coordinate measurement precision. Moreover, the a posteriori variance factor, when considering the time delay, is smaller than that while ignoring the time delay. This is an indication of slight improvement in the back-projection errors when considering the time delay. Table 5 shows the correlation matrix produced from the direct approach for the nHS-70
Table 6. Relative accuracy evaluated through the mean standard deviation (SD) of checkpoints derived from the bundle adjustment (BA) and the absolute accuracy evaluated through the mean/SD/root mean square error (RMSE) of the differences between the BA-based and surveyed coordinates of the five checkpoints from direct approach.

<table>
<thead>
<tr>
<th>Sensor and Date</th>
<th>Parameter</th>
<th>Relative Accuracy</th>
<th>Absolute Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean STD XYZ (m)</td>
<td>Mean XYZ (m)</td>
</tr>
<tr>
<td>nHS-70</td>
<td>9 July</td>
<td>Estimate Δt</td>
<td>−0.018 ± 0.011</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ignore Δt</td>
<td>−0.017 ± 0.010</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE Difference</td>
<td>0.001 ± 0.001</td>
</tr>
<tr>
<td></td>
<td>14 August</td>
<td>Estimate Δt</td>
<td>−0.007 ± 0.007</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ignore Δt</td>
<td>−0.005 ± 0.007</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE Difference</td>
<td>0.000 ± 0.002</td>
</tr>
<tr>
<td>uVS-307</td>
<td>9 July</td>
<td>Estimate Δt</td>
<td>−0.020 ± 0.010</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ignore Δt</td>
<td>−0.018 ± 0.003</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE Difference</td>
<td>0.001 ± 0.002</td>
</tr>
<tr>
<td></td>
<td>14 August</td>
<td>Estimate Δt</td>
<td>−0.006 ± 0.006</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ignore Δt</td>
<td>−0.006 ± 0.006</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE Difference</td>
<td>0.001 ± 0.000</td>
</tr>
<tr>
<td>nHS-199</td>
<td>29 May</td>
<td>Estimate Δt</td>
<td>0.007 ± 0.003</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ignore Δt</td>
<td>0.008 ± 0.002</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE Difference</td>
<td>0.001 ± 0.000</td>
</tr>
<tr>
<td></td>
<td>11 July</td>
<td>Estimate Δt</td>
<td>0.008 ± 0.007</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ignore Δt</td>
<td>0.011 ± 0.009</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE Difference</td>
<td>0.004 ± 0.003</td>
</tr>
</tbody>
</table>

*For time delay estimated, ignored.

Table 7. Estimated system parameters including the standard deviation and square root of a posteriori variance factor with a 200-ms artificial time delay introduced.

<table>
<thead>
<tr>
<th>Sensor and Date</th>
<th>Δt (ms)</th>
<th>ΔX (m)</th>
<th>ΔY (m)</th>
<th>Δα (°)</th>
<th>Δφ (°)</th>
<th>Δκ (°)</th>
<th>σ_a (Pixel)</th>
</tr>
</thead>
<tbody>
<tr>
<td>nHS-70</td>
<td>9 July</td>
<td>−0.60 ± 0.158</td>
<td>0.060 ± 0.334</td>
<td>−3.215 ± 0.343</td>
<td>179.751 ± 0.620</td>
<td>17.094</td>
<td></td>
</tr>
<tr>
<td></td>
<td>14 August</td>
<td>0.138 ± 0.230</td>
<td>179.416 ± 0.440</td>
<td>9.569</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nHS-199</td>
<td>29 May</td>
<td>0.112 ± 0.164</td>
<td>0.617 ± 0.340</td>
<td>−3.587 ± 0.334</td>
<td>−0.839 ± 0.616</td>
<td>9.569</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11 July</td>
<td>0.112 ± 0.196</td>
<td>0.412 ± 0.443</td>
<td>−5.046 ± 0.451</td>
<td>−0.408 ± 0.960</td>
<td>18.375</td>
<td></td>
</tr>
</tbody>
</table>

Objective 2: Feasibility of Using the Direct Approach with Large, Artificial Time Delay for Pushbroom Scanner Systems

The next section presents results for the direct approach while introducing a 200-ms artificial time delay for the nHS-70 and nHS-199 systems. The purpose of introducing this artificial time delay is to analyze the direct approach’s performance when having a large time delay in pushbroom scanner systems. The estimated parameters while considering/ignoring time delay for the nHS-70 and nHS-199 are displayed in Table 7. The time delays estimated for the 9 July and 14 August data sets for the nHS-70 system without the artificial time delay are 5.91 and 10.46 ms, respectively (shown in Table 4). The time delays estimated for those dates with a 200-ms artificial time delay are −194.09 and −189.54 ms, respectively. The difference between these estimates is exactly equivalent to the introduced 200-ms time delay. Similar results can be observed for the calibration parameters for the nHS-199 system by comparing Tables 4 and 7. These results show that the direct approach accurately estimates the time delay regardless of its magnitude. Furthermore, for all data sets ignoring the temporal calibration after introducing the 200-ms artificial time delay show a very large increase (and further from the nominal values) in the estimated lever arm
Table 8. Relative accuracy evaluated through the mean standard deviation (SD) of checkpoints derived from the bundle adjustment (BA) and the absolute accuracy evaluated through the mean/SD/root mean square error (RMSE) of the differences between BA-based and surveyed coordinates of the five checkpoints from direct approach while introducing a 200-ms artificial time delay.

<table>
<thead>
<tr>
<th>Sensor and Date</th>
<th>Parameter</th>
<th>Relative Accuracy</th>
<th>Absolute Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean STD XYZ (m)</td>
<td>Mean XYZ (m)</td>
</tr>
<tr>
<td>nHS-70</td>
<td>9 July</td>
<td>Estimate Δt</td>
<td>0.004 0.004 0.022</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ignore Δt</td>
<td>0.114 0.115 0.568</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE Difference</td>
<td></td>
</tr>
<tr>
<td></td>
<td>14 August</td>
<td>Estimate Δt</td>
<td>0.005 0.005 0.030</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ignore Δt</td>
<td>0.064 0.065 0.380</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE Difference</td>
<td></td>
</tr>
<tr>
<td>nHS-199</td>
<td>29 May</td>
<td>Estimate Δt</td>
<td>0.007 0.007 0.041</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ignore Δt</td>
<td>0.111 0.111 0.581</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE Difference</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11 July</td>
<td>Estimate Δt</td>
<td>0.008 0.007 0.041</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ignore Δt</td>
<td>0.062 0.013 0.194</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE Difference</td>
<td></td>
</tr>
</tbody>
</table>

For time delay estimated, ignored.

Table 9. Estimated system parameters including the standard deviation and square root of a posteriori variance factor for nHS-70 data sets from indirect approach while using 40-m flight lines only.

<table>
<thead>
<tr>
<th>Nominal Values</th>
<th>Δt (ms)</th>
<th>ΔX (m)</th>
<th>ΔY (m)</th>
<th>Δω (°)</th>
<th>Δφ (°)</th>
<th>Δκ (°)</th>
<th>δ̂ (Pixel)</th>
</tr>
</thead>
<tbody>
<tr>
<td>nHS-70</td>
<td>0</td>
<td>−0.02</td>
<td>0.06</td>
<td>0</td>
<td>0</td>
<td>180</td>
<td>N/A</td>
</tr>
<tr>
<td>Estimated System Parameters from Indirect Approach</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 July</td>
<td>N/A</td>
<td>1.667±0.845</td>
<td>0.06 (fixed)</td>
<td>−0.050±0.283</td>
<td>−2.297±1.189</td>
<td>179.93±0.040</td>
<td>0.789</td>
</tr>
<tr>
<td>14 August</td>
<td>N/A</td>
<td>−1.612±1.025</td>
<td>0.06 (fixed)</td>
<td>1.049±0.609</td>
<td>2.291±1.438</td>
<td>−179.95±0.054</td>
<td>0.881</td>
</tr>
</tbody>
</table>

Table 10. Correlation matrix of system parameters for 9 July nHS-70 indirect approach results (high correlation values, larger than 0.85, are in bold).

<table>
<thead>
<tr>
<th></th>
<th>ΔX</th>
<th>Δω</th>
<th>Δφ</th>
<th>Δκ</th>
</tr>
</thead>
<tbody>
<tr>
<td>ΔX</td>
<td>1</td>
<td>0.316</td>
<td>1.000</td>
<td>0.055</td>
</tr>
<tr>
<td>Δω</td>
<td>0.316</td>
<td>1</td>
<td>0.316</td>
<td>0.232</td>
</tr>
<tr>
<td>Δφ</td>
<td>−1.000</td>
<td>−0.316</td>
<td>1</td>
<td>0.055</td>
</tr>
<tr>
<td>Δκ</td>
<td>0.055</td>
<td>0.232</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

component in the flying direction (ΔX lever arm). In terms of δ̂ values, results from the direct approach before/after introducing a large artificial time delay are identical (i.e., within 1 pixel). However, as shown in Table 7, the δ̂ values after estimating system parameters while ignoring time delay are in the range of 9 to 18 pixels. This suggests, as expected, that the spatial calibration is not capable of accurately estimating the mounting parameters when the system has a large time delay (thus leading to large back-projection errors).

Table 8 lists the relative accuracy—as evaluated by the mean of the BA-based standard deviations for the checkpoints—and the absolute accuracy—as evaluated by comparing the ground coordinates of the checkpoints to surveyed ground coordinates. The differences in the absolute accuracy (RMSE values) while considering/ignoring time delay are much larger now. When the artificial time delay of 200 ms was introduced, there is an overall horizontal and vertical improvement while considering time delay. The improvement in the horizontal component of the RMSE while using the direct approach to estimate time delay is in the range of 4 to 14 cm, which is approximately two to eight times that of the GSD of the system at 20-m flying height. The vertical component of the RMSE shows an even larger improvement by using the direct approach for estimating time delay, with as much as a 105.0-cm difference. This shows a significant improvement in the absolute accuracy from the direct approach in the presence of a large time delay. Furthermore, even though there was a large time delay, the RMSE measures are at the same level when having small synchronization errors. For example, the RMSE XYZ components while conducting spatial and temporal calibration for the nHS-199 11 July data set without the artificial time delay, shown in Table 6, are identical to the RMSE XYZ components with the artificial time delay introduced, shown in Table 8. Besides the time delay, other estimated system parameters when introducing the artificial time delay are also identical to the ones without artificial time delay. This shows that the direct approach accurately estimates the time delay and achieves similar absolute accuracy regardless of whether the time delay in the system is minimal or significant.

**Objective 3: Feasibility of Using the Indirect Approach to Estimate Time Delay for Pushbroom Scanner Systems**

The feasibility of the indirect approach to estimate time delay for pushbroom scanner systems is discussed in this section. The objective was tested on the nHS-70. One should recall from the bias impact analysis that the indirect approach requires a single flying speed (which further leads to a single flying height for pushbroom scanners) so that the correlation between the lever arm in the flying direction and time delay is exploited. The estimated parameters for the indirect approach for the nHS-70 9 July and 14 August 40-m flight lines are presented in Table 9. It is worth mentioning that the lever arm component across the flight direction (i.e., ΔY) is fixed due to the high correlation between ΔY and Δω at a single flying height. As can be seen in Table 9, the lever arm component in the flying direction (i.e., ΔX) for the two tests are extremely large compared to the nominal values. This is expected since the time delay is being absorbed by such a component. The reported standard deviations for the estimated parameters are extremely large (in the ±1 m/±1° range). The correlation matrix of system parameters for the 9 July data set is shown in Table 10. Here, a problem was identified with ΔX and Δφ being 100% correlated (highlighted in bold). Given the high
correlation as well as large standard deviations, system parameters are not reliable, and step 2 of the indirect approach was not performed. The bias impact analysis revealed that this correlation was expected, and changes in flying height are needed to decouple the lever arm components and boresight angles. Therefore, the indirect approach should not be used for pushbroom scanners. The direct approach is the recommended one for estimating time delay during system calibration for pushbroom scanner systems.

Conclusions and Recommendations for Future Research

UAVs equipped with GNSS/INS-assisted imaging systems are often used for accurate 3D reconstruction. Spatial and temporal system calibration is crucial for deriving accurate 3D spatial information. In this article, the direct and indirect system calibration approaches were presented and analyzed for pushbroom scanner systems. Optimal flight configuration was derived through a bias impact analysis for pushbroom scanner systems. It was determined through the bias impact analysis that the indirect approach was not appropriate due to the small AFOV of pushbroom scanner systems and the single flying height/speed requirements of that approach. This hypothesis was then confirmed through the experimental results. The direct approach results for all systems used in this study showed that the time delay can be accurately estimated for both small and large synchronization errors. The results show that even when there is a relatively small synchronization error, ignoring the time delay could lead to poor reconstruction, especially when dealing with unfavorable distribution of tie points in the acquired scenes. Therefore, it is suggested to always consider spatial and temporal calibration even if a minimal time delay is suspected. An added advantage of the proposed direct approach is eliminating the need for GCPs provided that an optimal configuration for the data acquisition is maintained.

Future work will focus on automating the tie point measurement process. In this study, the results show that tie point distribution is extremely important for deriving accurate estimates of system parameters. It can be difficult to ensure both a large number and a good distribution while using manually identified tie points. Therefore, future work will concentrate on both automated extraction of tie points from frame and pushbroom scanner imagery as well as incorporating those tie points into the direct system calibration approach. Also, future work will focus on verifying the direct approach for pushbroom scanner sensors onboard a variety of satellite and airborne platforms.

Acknowledgments

The information, data, or work presented herein were funded by the Advanced Research Projects Agency Energy (ARPAE), U.S. Department of Energy, under award no. DE-AR0001135. The views and opinions of the authors expressed herein do not necessarily state or reflect those of the U.S. government or any agency thereof.

References


Least Squares Adjustment with a Rank-Deficient Weight Matrix and Its Applicability to Image/Lidar Data Processing

Radhika Ravi and Ayman Habib

Abstract
This article proposes a solution to special least squares adjustment (LSA) models with a rank-deficient weight matrix, which are commonly encountered in geomatics. The two sources of rank deficiency in weight matrices are discussed: naturally occurring due to the inherent characteristics of LSA mathematical models and artificially induced to eliminate nuisance parameters from LSA estimation. The physical interpretation of the sources of rank deficiency is demonstrated using a case study to solve the problem of 3D line fitting, which is often encountered in geomatics but has not been addressed fully to date. Finally, some geomatics-related applications—mobile lidar system calibration, point cloud registration, and single-photo resection—are discussed along with respective experimental results, to emphasize the need to assess LSA models and their weight matrices to draw inferences regarding the effective contribution of observations. The discussion and results demonstrate the vast applications of this research in geomatics as well as other engineering domains.

Introduction
Least squares adjustment (LSA) is used to solve overdetermined systems of equations based on minimizing the squared sum of weighted observation residuals. LSA is a commonly used tool for solving a multitude of problems throughout the field of engineering and beyond. The field of geomatics relies heavily on LSA for a vast range of activities, including leveling, bundle adjustment, global navigation satellite system positioning, and so on. It was first developed in the first half of the 19th century by Gauss (1823). Apart from LSA, other methods of adjustment in surveying have also been developed, such as the Kalman filter (Kalman 1960), least squares collocation (Moritz 1978) and total least squares (TLS; Van Huffel and Vandewalle 1991; Dogan and Altan 2010).

The focus of this article is to introduce an extension of the concept of LSA to scenarios that might result in a rank-deficient weight matrix. While existing LSA mathematical models can address most engineering problems, those pertaining to the field of geomatics are highly prone to giving rise to scenarios involving a rank-deficient weight matrix, which causes complications in finding solutions based on existing LSA models. A rank-deficient weight matrix is introduced in the model in problems that are as simple as line or plane fitting for a given set of 3D points. Such LSA models propagate to a multitude of problems in the fields of lidar and photogrammetry that rely on such 3D line/plane fitting for processing and extracting information from lidar and image-based 3D point clouds. The applicability of LSA with a rank-deficient weight matrix to such a vast range of activities related to the field of geomatics serves as the motivation for this article, which aims to propose an approach to solve such LSA models.

This article starts with a discussion of prior research related to the field of LSA, along with a brief summary of widely used LSA models—namely, the Gauss–Markov and Gauss–Helmert models—which form the basis of the methodology proposed in this article. Then the next section steers toward the introduction of an LSA formulation pertaining to models with a rank-deficient weight matrix. Rank deficiency is discussed in terms of its mathematical and physical interpretation based on its source: naturally arising rank deficiency as a result of the mathematical model for LSA and artificially induced rank deficiency targeting the removal of nuisance parameters during LSA. Based on the generic mathematical formulation, the next section discusses a case study of 3D line fitting to provide conclusive approaches for line parameters estimation using naturally and artificially induced rank deficiency in the weight matrix and to demonstrate the validity of the proposed solution to such LSA models. The next section illustrates the manifestation of a rank-deficient weight matrix in LSA models used in some of the most commonly encountered applications in the field of geomatics: 3D plane fitting, mobile lidar system calibration, feature-based registration of overlapping point clouds, and single-photo resection. The final section summarizes the findings of this research.

Related Work
The most commonly used model for LSA is the Gauss–Markov model, which states that the ordinary least squares estimator results in the lowest sampling variance within the class of linear unbiased estimators (Henderson 1975). However, in situations where the design matrix for LSA is ill-conditioned, the original least squares estimator is not optimal. For such cases, a biased least squares estimator is used instead of the original one. Several types of biased estimators have been proposed over the years: ordinary ridge, principal components, combined principal components, single-parametric principal components, and root-root (Massy 1965; Hoerl and Kennard 1970; Vinod and Ullah 1981; Xia 1988; Hu 1990). The biased estimators are then extended to the Gauss–Markov model, as proposed by Gui and Liu (2000), with the choice of biased estimator depending on the application for which LSA is being used.

Contributed by Jose M. Peña, September 1, 2020 (sent for review March 25, 2021; reviewed by Linfu Xie, Devrim Akca).
The Gauss–Markov model can address problems where the dependent variables (or observations) are a linear function of unknown parameters in the model. On the other hand, more generic cases involving nonlinear functions in both observations and parameters are transformed into a linear system that is dealt with by the mixed model, also known as the Gauss–Helmert model (Helmert 1907). Furthermore, Wentworth (1965) introduced a rigorous LSA method to be applied to nonlinear equations. Mikhail and Ackermann (1982) addressed the different LSA models, including the observations-only model, indirect observations model, and the mixed model (Gauss–Helmert model) combining observations and parameters. The application of LSA to the field of geomatics has been explored by Hirvonen (1971), Triggs et al. (2000), and Mikhail et al. (2001). Including activities such as leveling and bundle adjustment. While ordinary LSA deals with cases where the variance-covariance matrix for errors is identity, the extension of ordinary LSA known as weighted LSA deals with problems where the variance-covariance matrix for the observations is different from an identity matrix, with unequal diagonal elements and nonzero nondiagonal elements (Mandel 2012). Markovsky et al. (2006) started by introducing an element-wise weighted total LSA model to generalize the total least squares estimator for cases with increasingly generic noise assumptions. However, this generalization increased the computational difficulty of finding a solution to such problems, whereas total least squares models tend to have closed-form solutions. Significant research has been conducted to propose iterative algorithms to derive the solution to weighted total least squares models (Schafrin and Wieser 2008; Shen et al. 2011). Fang (2013) discussed the necessary and sufficient conditions for a weighted least squares model with fixed and random parameters. Further, Fang (2015) addressed the weighted total least squares problems with constraints.

As already stated, while a great deal of research has been conducted for cases where the normal equation matrix is rank-deficient, there is no insight provided for cases where the weight matrix of the errors or observation equations is rank-deficient. Such a rank deficiency might arise either naturally or artificially. Whereas a naturally rank-deficient weight matrix is indicative of the nature of the LSA mathematical model, an artificial rank reduction may be induced to address the elimination of certain parameters from the LSA estimation model. Commonly occurring scenarios that give rise to such models are well-known problems dealing with 3D line and plane fitting to a given set of points. Moreover, 3D feature-fitting problems serve as critical stepping-stones toward various lidar and photogrammetric activities, such as mobile lidar system calibration, bundle adjustment, registration of overlapping point clouds using linear/planar features, and so on. This serves as the motivation for the research proposed in this article addressing the case for LSA involving a rank-deficient weight matrix, which is further applied to a case study pertaining to 3D line fitting.

**Review: Gauss–Markov Model**

For any linear system, the mathematical model for LSA can be written as in the following equation, which is the most commonly used model (Mikhail and Ackermann 1982), also known as the Gauss–Markov model:

\[
y = Ax + e; \quad e \sim (0, \sigma_e^2 P^{-1})
\]

where \( y \) denotes the \( n \times 1 \) vector of observations with an associated \( n \times n \) weight matrix denoted by \( P \); \( A \) is the \( n \times m \) design matrix; \( x \) is the \( m \times 1 \) vector of unknown parameters (with \( n > m \)); and \( \sigma_e^2 \) is the a priori variance factor. The corresponding error term \( e \) is given in

\[
e = y - Ax
\]

The objective of LSA is to minimize \( e^T P e \), which leads to the least squares estimate of \( x \) (or \( \hat{x} \)) given by

\[
\hat{x} = (A^T PA)^{-1} A^T Py
\]

The predicted residuals \( \hat{e} \) are given by

\[
\hat{e} = y - A\hat{x}
\]

The redundancy for the Gauss–Markov LSA model is defined as the difference between the number of observations \( n \) and the number of unknown parameters \( m \), which results in the a posteriori variance factor \( \hat{\sigma}_e^2 \):

\[
\hat{\sigma}_e^2 = \frac{\hat{e}^T P \hat{e}}{n-m}
\]

Further, the variance-covariance matrix (or dispersion matrix) \( \Sigma(\hat{x}) \) of the estimated unknown parameters is given by

\[
\Sigma(\hat{x}) = \hat{\sigma}_e^2 (A^T PA)^{-1}
\]

**Review: Gauss–Helmert Model**

A typical nonlinear function involving both observations and unknown parameters is represented by the following equation:

\[
f(Y - e, X) = 0
\]

where \( Y \) is the \( n \times 1 \) vector of observations, \( X \) is the \( m \times 1 \) vector of unknown parameters, and \( e \) is the \( n \times 1 \) vector of random errors contaminating the observations. This model can be linearized by applying Taylor series expansion and ignoring second- and higher-order terms. The resultant linearized model is given in the next equation, where \( X_0 \) is an initial guess of the unknown parameters:

\[
f(Y, X_0) + \frac{\partial f}{\partial Y} \bigg|_{Y,X_0} (-e) + \frac{\partial f}{\partial X} \bigg|_{Y,X_0} dX = 0
\]

Equation 8 can be rewritten in the form given in the following equation, which depicts a mathematical model with \( n \) observation equations, which may be less than the total number of observations \( n \):

\[
y = Ax + Be; \quad e \sim (0, \sigma_e^2 P^{-1})
\]

where \( y = f( Y, X_0) \) is the new \( n \times 1 \) vector of observation equations (where \( n_s \leq n \)), \( A = -\frac{\partial f}{\partial X} \bigg|_{Y,X_0} \) is the \( n_s \times m \) design matrix composed of the partial derivatives with respect to the unknown parameters, and \( B = \frac{\partial f}{\partial Y} \bigg|_{Y,X_0} \) is the \( n_s \times n \) matrix composed of partial derivatives with respect to the observations. The mathematical model in Equation 9 is also widely known as the Gauss–Helmert model (Mikhail and Ackermann 1982). Defining \( \sigma = Be \) as the combined error term, the
Gauss–Helmert model can be transformed to a form similar to the Gauss–Markov model:

\[ y = Ax + \bar{e}; \quad \bar{e} \sim (0, \sigma_0^2BP^{-1}B^T) \]  \hspace{1cm} (10)

As in the Gauss–Markov model, the LSA aims at estimating the unknown parameters that minimize the sum of squared weighted residuals—that is, the estimates of unknown parameters that minimize \( \bar{e}^T (BP^{-1}B^T) \bar{e} \). The least squares estimate of \( x \) in the Gauss–Helmert model is

\[ \hat{x} = (A^T (BP^{-1}B^T)^{-1} A)^{-1} \left( A^T (BP^{-1}B^T)^{-1} y \right) \]  \hspace{1cm} (11)

and the predicted residuals \( \hat{e} \) are given by

\[ \hat{e} = y - A\hat{x} \]  \hspace{1cm} (12)

The redundancy in the Gauss–Helmert LSA model is defined as the difference between the number of equations \( n_e \) and the number of unknown parameters \( m \), which results in the \( a \) posteriori variance factor \( \hat{\sigma}_0^2 \):

\[ \hat{\sigma}_0^2 = \frac{\hat{e}^T (BP^{-1}B^T)^{-1} \hat{e}}{n_e - m} \]  \hspace{1cm} (13)

Further, the variance-covariance matrix (or dispersion matrix) \( \Sigma(\hat{x}) \) of the estimated unknown parameters is given by

\[ \hat{\Sigma}(\hat{x}) = \hat{\sigma}_0^2 \left( A^T (BP^{-1}B^T)^{-1} A \right)^{-1} \]  \hspace{1cm} (14)

**LSA Model with a Rank-Deficient Weight Matrix**

It can be seen that the least squares estimates for the unknown parameters in the Gauss–Helmert model can be derived using the foregoing equations only when the matrix \( BP^{-1}B^T \) is invertible or, in other words, a full-rank matrix. However, for some applications related to lidar and photogrammetry, as mentioned earlier, it may be rank-deficient. This section discusses in detail the mathematical formulation of the solution to LSA models with a rank-deficient weight matrix. Rank deficiency can be introduced in two ways—naturally arising rank deficiency owing to the mathematical LSA model or artificially induced rank deficiency to eliminate nuisance parameters from the LSA model—both of which are discussed in detail in the following subsections.

**Naturally Rank-Deficient Weight Matrix**

Keeping in mind the Gauss–Helmert model, it can be inferred that the source of rank deficiency would be the matrix \( B \) (composed of the partial derivatives of the nonlinear function with respect to the observations), since the matrix \( P \) denoting the weight of observations can be safely assumed to be of full rank. Such rank deficiency owing to the natural form of the matrix of partial derivatives with respect to the observations is regarded in this section as naturally arising rank deficiency, which is mainly encountered when dealing with Gauss–Helmert models, since matrix \( B \) is responsible for such rank deficiency. It will be shown later that a naturally arising rank deficiency of the variance-covariance/weight matrix of the combined error term in the Gauss–Helmert LSA model targets the elimination of error terms whose minimization does not contribute to the estimation of the unknown parameters. In other words, such rank deficiency can be interpreted as an indication of the effective contribution of observations to redundancy—that is, the solution of the unknown parameters at hand.

In this case, the rank deficiency of the variance-covariance matrix \( BP^{-1}B^T \) implies that it is noninvertible, and thus the resultant weight matrix of the combined error term is obtained by evaluating the Moore–Penrose pseudo-inverse of the matrix (Penrose 1955), denoted by a superscript plus sign. Based on the mathematical formulation of generic LSA models with a rank-deficient weight matrix presented in the Appendix, the least squares estimate of the unknown parameters in this case are given by

\[ \hat{x} = (A^T (BP^{-1}B^T)^+ A)^{-1} A^T (BP^{-1}B^T)^+ y \]  \hspace{1cm} (15)

The predicted residuals \( \hat{e} \) are computed as mentioned earlier in Equation 12. From the Appendix, the redundancy for the Gauss–Helmert LSA model with a rank-deficient weight matrix will now be defined as the difference between the rank of the resultant weight matrix of the combined error term and the number of unknown parameters \( m \), which results in the \( a \) posteriori variance factor \( \hat{\sigma}_0^2 \):

\[ \hat{\sigma}_0^2 = \frac{\hat{e}^T (BP^{-1}B^T)^+ \hat{e}}{\text{rank}(BP^{-1}B^T) - m} \]  \hspace{1cm} (16)

Further, the variance-covariance matrix (or dispersion matrix) \( \hat{\Sigma}(\hat{x}) \) of the estimated unknown parameters is given by

\[ \hat{\Sigma}(\hat{x}) = \hat{\sigma}_0^2 \left( A^T (BP^{-1}B^T)^+ A \right)^{-1} \]  \hspace{1cm} (17)

This discussion provides vital insight about dealing with LSA models in the field of geomatics, where the usual approach adopted to gauge the effective contribution of observations to redundancy is to conduct a careful analysis of the geometric contributions of equations/observations to the estimation of unknowns. However, such an approach requires a thorough understanding of the problem addressed by the LSA model in question. A geometric interpretation of problems is not always straightforward, especially when the problems are not intuitive enough. However, the mathematical approach proposed here can reliably deduce the effective contribution by conducting a proper error propagation and investigating the rank of the resultant weight matrix associated with the relevant LSA.

**Artificially Induced Rank-Deficient Weight Matrix**

In this subsection, we present an alternate approach to estimating unknown parameters by inducing an artificially modified, rank-deficient weight matrix. Artificially introducing a rank-deficient weight matrix targets the elimination of nuisance parameters from the LSA model, unlike the naturally rank-deficient weight matrix discussed earlier, which addresses elimination of the error terms that do not contribute to redundancy—that is, do not contribute to the solution to the problem at hand.

Say we start with a model analogous to the Gauss–Markov model, as given by the following equation, which includes a nonrandom vector \( d \) composed of additional unknown parameters (or nuisance parameters):

\[ y = Ax + d + e; \quad e \sim (0, \sigma_0^2P^{-1}) \]  \hspace{1cm} (18)
In this case, $x$ denotes the unknown parameters whose estimates are desired; the corresponding target function for LSA is given by

$$
\phi(x, d) = e^T P e = (y - Ax - d)^T P(y - Ax - d) = \min_{x, d}
$$

(19)

The elimination of nuisance parameters from the LSA estimation model can be achieved by artificially modifying the weight matrix $P$ to obtain $P'$ (modified weight matrix), which can nullify the nonrandom vector $d$—that is, $P' d = 0$. The objective function for LSA is then modified to the following form:

$$
\phi(x) = e^T P' e = (y - Ax)^T P'(y - Ax) = \min_{x, d}
$$

(20)

(21)

Based on the generic discussion in the Appendix of LSA models with a rank-deficient weight matrix, the resultant least squares estimate of unknown parameters $\hat{x}$ is given by

$$
\hat{x} = (A^T P'A)^{-1} A^T P' y
$$

The predicted residuals $\hat{e}$ are given by

$$
\hat{e} = y - A\hat{x} - d
$$

(22)

From the Appendix, the redundancy of this LSA model is given by the difference between the rank of the modified weight matrix and the number of unknown parameters $m$, which results in the a posteriori variance factor $\hat{\sigma}_0^2$:

$$
\hat{\sigma}_0^2 = \frac{\hat{e}^T P' \hat{e}}{\text{rank}(P') - m} = \frac{(y - A\hat{x})^T P'(y - A\hat{x})}{\text{rank}(P') - m}
$$

(23)

Note that computing the a posteriori variance factor does not require any knowledge of the nonrandom vector $d$ in Equation 22, since the rank-deficient weight matrix would nullify that vector. So the a posteriori variance factor can be derived using only the observations and estimated unknowns. Furthermore, the variance-covariance matrix (or dispersion matrix) $\Sigma(\hat{x})$ of the estimated unknown parameters is given by

$$
\Sigma(\hat{x}) = \hat{\sigma}_0^2 (A^T P'A)^{-1}
$$

(24)

In summary, an artificially induced rank-deficient weight matrix eliminates nuisance parameters from the LSA problem, thus allowing for more general approaches to common problems in geomatics, such as mobile lidar system calibration, registration of overlapping 3D point clouds, and single-photo resection (discussed later).

**Case Study: 3D Line Fitting**

Having discussed the solution and interpretation of generic LSA models involving a naturally occurring or artificially induced rank-deficient weight matrix, we now discuss the application of such LSA models to the problem of 3D line fitting. Line fitting to 2D points has been optimally solved by Pearson (1901). The scope of the TLS solution was also explored, with respect to its applicability to 2D line fitting, by Schaffrin et al. (2006). However, line fitting in 3D space still remains unresolved, since it presents more challenges. The representation scheme or parameters involved in the case of 3D lines was addressed by Plücker (1852), who indicated that a 3D line can be defined using four parameters. A TLS adjustment method was explored by Snow and Schaffrin (2016), who discussed the minimal parametrization of 3D lines and a TLS-based approach to estimating the four line parameters. In this section, the mathematical formulation of an LSA model with a rank-deficient weight matrix will be applied to 3D line fitting to demonstrate the applicability of such LSA models and to provide deeper insights into their interpretation and usage.

The objective function of an LSA model for 3D line fitting is to minimize the squared sum of normal distances from the fitted line to the given 3D points. Using Figure 1 as a reference schematic illustration, let us assume that a candidate line is defined by a point $a$ on the line with coordinates $(x_0, y_0, z_0)$ and a unit vector along the line direction, $(u_x, u_y, u_z)$. For any observed point $b$ whose coordinates are $(x_i, y_i, z_i)$, its projection onto the given line is denoted as $b_i(x_i, y_i, z_i)$ and the normal distance of the point from the line is denoted as $(d_x, d_y, d_z)$. The noise-free 3D coordinates for the point $b$ are denoted by $(\bar{x}, \bar{y}, \bar{z})$ as given in the following equation, where $(e_x, e_y, e_z)$ denotes the random errors contaminating the 3D point coordinates:

$$
\begin{align*}
\bar{x} &= x_i - e_x \\
\bar{y} &= y_i - e_y \\
\bar{z} &= z_i - e_z
\end{align*}
$$

(25)

Figure 1. Schematic illustration of 3D line fitting by minimizing the normal distances between the points and the fitted line: $(u_x, u_y, u_z)$ denotes the unit vector along the line direction, $a(x_0, y_0, z_0)$ denotes a point on the line, $b(x_i, y_i, z_i)$ denotes an observed point, $b_i(x_i, y_i, z_i)$ denotes the projection of the observed point on the line, and $(d_x, d_y, d_z)$ denotes the normal distance of the observed point from the line.

**3D Line Fitting Involving a Naturally Rank-Deficient Weight Matrix**

This subsection aims to assess the effective contribution of a set of observations to the determination of unknown 3D line parameters. This is achieved by formulating the linearized LSA model for 3D line fitting, followed by analyzing and interpreting the resultant weight matrix for the error terms coming from the observation equations. This approach results in an LSA model with a naturally rank-deficient weight matrix, and it will be shown that such a matrix eliminates random errors that do not contribute to the estimation of line parameters. More specifically, it will be shown that the resultant rank-deficient weight matrix eliminates the random error component lying along the direction of the line parameters.

The vector $\mathbf{b}_n \mathbf{b}$ from point $b$ to the line can be expressed as

$$
\mathbf{b}_n \mathbf{b} = \mathbf{ab} - \mathbf{ab}_n
$$

(26)

where $\mathbf{ab}$ and $\mathbf{ab}_n$ are given by
To assess the effective contribution of a set of observations, an in-depth analysis of the weight matrix associated with the combined error term is conducted. This weight matrix can be obtained as the inverse of the variance-covariance matrix. But note that the matrix $\Sigma_e$ is rank-deficient with rank 2; this implies that the variance-covariance matrix of the combined error term $\sigma_e^2$ is noninvertible. Because of the rank deficiency of the variance-covariance matrix $\Sigma_e$, the desired weight matrix of the combined error term is obtained by computing the Moore–Penrose pseudo-inverse of $\Sigma_e$. The first step to finding the Moore–Penrose pseudo-inverse is to conduct an eigenvalue decomposition of the matrix $B_i$. Taking into consideration that $B_i$ is a positive semidefinite symmetric matrix, its eigenvalue decomposition can be written as

$$B_i = V \Lambda V^T \tag{34}$$

where $V$ is the matrix composed of the eigenvectors of $B_i$ and $\Lambda$ is a diagonal matrix with the corresponding eigenvalues. The eigenvalue decomposition of $B_i$ indicates that its eigenvalues would be $\lambda_1 = 0$ and $\lambda_2 = \lambda_3 = 1$. The eigenvector corresponding to $\lambda_1$ is the vector $[u_x, u_y, u_z]^T$, which lies along the direction of the line. Since the eigenvectors of a symmetric matrix form a set of orthonormal basis vectors (Osnaga 2005), the remaining two eigenvectors will correspond to the two directions $v$ and $w$ that are normal to the line, as shown in Figure 2. The derived eigenvectors of $B_i$ indicate that the matrix $V$ of eigenvectors actually represents the rotation matrix from the local 3D line coordinate system to the mapping frame coordinate system, or $R_{uvw}$ in Figure 2.

![Figure 2. Schematic illustration of the mapping frame and local 3D line coordinate systems.](image)

Analyzing the combined error term $\mathbf{v}_i$ (or $B_i \mathbf{e}$) by substituting the foregoing results for the eigenvalue decomposition of $B_i$ will result in

$$\bar{\mathbf{e}}_i = V \Lambda^T \mathbf{e} = R_{uvw}^T \Lambda R_{uvw} \mathbf{e}_i$$

$$e_i = R_{uvw}^T \mathbf{e}_i = R_{uvw}^T \begin{bmatrix} e_{i,v} \\ e_{i,u} \\ e_{i,w} \end{bmatrix} = R_{uvw}^T \mathbf{e}_i \tag{35}$$

This simplification indicates that the combined error term indeed nullifies the error $e_{i,v}$ along the line while only

$\bar{\mathbf{e}}_i$
considering the errors \( e_x \) and \( e_y \) in the two directions normal to the 3D line to conduct the LSA for 3D line fitting. This in turn indicates that the effective contribution from each set of observations arising from a 3D point is only two equations, as indicated by the rank of the matrix \( B_i \) and interpreted using the eigenvector matrix of \( B_i \) to consider only the errors along the two normal directions to the line.

Having obtained the eigenvalue decomposition of the matrix \( B_i \), the next step is to find the Moore–Penrose pseudoinverse of the variance-covariance matrix \( \Sigma_i \), as given by

\[
\Sigma_i^{-1} = \sigma_i^2 \left( B_i P_i^{-1} B_i^T \right)^T = \sigma_i^2 V A^T V^T = \sigma_i^2 V V^T = \sigma_i^2 V (1 \ 0 \ 0) \quad (36)
\]

Finally, the expression \( \tilde{e}_i^T \left( B_i P_i^{-1} B_i^T \right)^T \tilde{e}_i \) can be simplified as

\[
\tilde{e}_i^T \left( B_i P_i^{-1} B_i^T \right)^T \tilde{e}_i = \tilde{e}_i^T R_{xy}^i A \tilde{e}_i = \tilde{e}_i^T \Sigma_i^{-1} \tilde{e}_i
\]

which in turn, indicates that the objective function of the resultant LSA model in this approach for 3D line fitting works toward minimizing the normal distance between the line and the given set of 3D points. Thus, the LSA model developed in this section for 3D line fitting works toward estimating the line parameters by minimizing the squared sum of errors in the two directions normal to the line, and this is ensured by the naturally occurring rank-deficient weight matrix of the combined error term in the LSA model. So while it appears from initial assessment that each 3D point contributes three equations to the solution, the analysis of the rank-deficient weight matrix indicates an effective contribution of only two equations per point.

### 3D Line Fitting While Artificially Modifying the Weight Matrix

The previous subsection discussed the effective contribution of observations to 3D line fitting. In this subsection, the main goal is to propose an alternative two-step approach for 3D line fitting by eliminating nuisance parameters in the LSA model. This is achieved by introducing the concept of artificial weight modification, resulting in rank notation. Note that the previously proposed one-step approach can estimate 3D line parameters as accurately as the two-step approach presented in this section. However, this discussion uses the problem of 3D line fitting to demonstrate artificial manipulation of LSA models, which in turn lays the foundation for more general approaches to other common problems in geomatics such as system calibration, 3D point cloud registration, and single-photo resection. Starting with the mathematical model of the 3D line that has been discussed so far, any point on the line can be written as

\[
\begin{bmatrix}
x_x \\
y_x \\
z_x
\end{bmatrix} = \begin{bmatrix}
x_0 \\
y_0 \\
z_0
\end{bmatrix} + \begin{bmatrix}
u_x \\
u_y \\
u_z
\end{bmatrix} \cdot \begin{bmatrix}
e_x \\
e_y \\
e_z
\end{bmatrix} = \begin{bmatrix}
x_0, \sigma_0^2 P_{xy}^{-1} e_y
\end{bmatrix} (38)
\]

The 3D line fitting is done in a two-step process: estimating a point \( (x_0, y_0, z_0) \) on the line and then estimating the line direction vector \( (u_x, u_y, u_z) \) — not necessarily a unit vector. For the first step, we need to find the directional unknowns from the equation. Assuming that prior information regarding the orientation of the line is known, we can define a rotation matrix from the local 3D line coordinate system to the mapping frame coordinate system \( R_{xy} \), as shown in Figure 2. Note that the prior orientation information of a 3D line can be derived by conducting principal component analysis on the 3D point coordinates, where the resultant eigenvector corresponding to the largest eigenvalue would serve as the initial estimate of the line direction. A modified weight matrix in the local 3D line coordinate system \( (uvw) \), shown in Figure 2, is given by

\[
P_{uv}^w = \begin{bmatrix}
0 & 0 & 0 \\
0 & P_v & P_{uv} \\
0 & P_{uv} & P_w
\end{bmatrix}
\]

Here the weight along the line direction (or \( v \)-axis) is set to zero, while those in the directions normal to the line are retained. This modified weight matrix can be transformed into the mapping frame coordinate system with the help of the prior information of the line orientation, as given by

\[
P_{xy} = R_{uvw} P_{uv} P_{uvw}^{-1}
\]

Note that the modified weight matrix would nullify any vector component lying along the line, thus serving the purpose of the first step of line fitting by eliminating the line directional unknowns from the model. This claim can be mathematically proven by analyzing the impact of \( P_{xy} \) on a vector along the line given by \( s[u_x u_y u_z]^T \):

\[
P_{xy} S_i = \begin{bmatrix}
d_x \\
d_y \\
d_z
\end{bmatrix} = \begin{bmatrix}
0 & 0 & 0 \\
0 & P_v & P_{uv} \\
0 & P_{uv} & P_w
\end{bmatrix} \begin{bmatrix}
d_x \\
d_y \\
d_z
\end{bmatrix}
\]

\[
\Rightarrow P_{xy} S_i = \begin{bmatrix}
u_x \\
u_y \\
u_z
\end{bmatrix} = 0
\]

Here \( d \) denotes the component along the \( v \)-axis of the aforementioned vector along the line. Formulating the LSA model for 3D line fitting using the modified weight matrix results in

\[
\begin{bmatrix}
x_i \\
y_i \\
z_i
\end{bmatrix} = \begin{bmatrix}
x_0 \\
y_0 + s_i u_y + e_y \\
z_0 + e_z
\end{bmatrix}, \quad \text{where} \quad \begin{bmatrix}
e_x \\
e_y \\
e_z
\end{bmatrix} ~ (0, \sigma_0^2 P_{xy}^{-1}) (42)
\]

Using the property that the modified weight matrix nullifies any vector lying along the line direction, the solution to this LSA model is given by

\[
\begin{bmatrix}
\hat{x}_0 \\
\hat{y}_0 \\
\hat{z}_0
\end{bmatrix} = \left( \sum_i (A^T P_{xy} A)^{-1} \right)^{-1} \left( \sum_i (A^T P_{xy} A) \begin{bmatrix}
x_i \\
y_i \\
z_i
\end{bmatrix} \right)
\]

where \( A \) is a \( 3 \times 3 \) identity matrix, as gleaned from Equation 42. Note that the matrix \( \sum P_{xy}^{-1} \) will be rank-deficient, since an infinite number of points lie on the line. This rank
deficiency is addressed by eliminating one of the three parameters \(x_0, y_0,\) and \(z_0\) from the estimation model. In other words, one of the three coordinates is fixed and the other two are estimated. The parameter to be fixed is decided based on the following three scenarios:

1. For lines oriented mainly along the x-axis—that is, \((u_x > u_y)\) and \((u_x > u_z)\)—\(x_0\) is fixed at 0 and LSA solves only for \(y_0\) and \(z_0\).
2. For lines oriented mainly along the y-axis—that is, \((u_y > u_x)\) and \((u_y > u_z)\)—\(y_0\) is fixed at 0 and LSA solves only for \(x_0\) and \(z_0\).
3. For lines oriented mainly along the z-axis—that is, \((u_z > u_x)\) and \((u_z > u_y)\)—\(z_0\) is fixed at 0 and LSA solves only for \(x_0\) and \(y_0\).

Having estimated the point on the line, the next step is to estimate the line direction. The line equation is rearranged to give

\[
\begin{align*}
\frac{x_i - x_0}{u_x} &= s_1 \frac{u_x}{u_y} + e_{x_i} \\
\frac{y_i - y_0}{u_y} &= s_1 \frac{u_y}{u_x} + e_{y_i} \\
\frac{z_i - z_0}{u_z} &= s_1 \frac{u_z}{u_x} + e_{z_i}
\end{align*}
\]

In this step, the unknown parameters to be estimated are \(s_1, u_x, u_y,\) and \(u_z,\) where the vector \((u_x, u_y, u_z)\) is not constrained to be a unit vector. Similar to the first step, to avoid singularity in the inversion of the resultant rank-deficient normal equation matrix, one of the directional parameters needs to be fixed while the other two are solved for. The parameter to be fixed is chosen based on one of the three scenarios, determined from the prior information of the line direction. The direction parameter corresponding to the major axis of spread of the line is fixed to unity.

These two steps conclude the approach to 3D line fitting based on the modified weight matrix. Note that each of these steps will have an associated \(a\) posteriori variance factor for LSA, which is computed based on the redundancy in each step. In this case, the rank of the modified weight matrix determines the redundancy for the first step, resulting in a redundancy of \((2n - 2)\), where \(n\) denotes the total number of 3D points along the line. For the second step, the total number of observations is \(3n\) and the total number of unknown parameters includes the \(n\) scaling factors and two directional parameters, resulting in a redundancy of \((2n - 2)\).

**Simulation and Experimental Results**

Having presented a detailed description of a case study involving LSA with a rank-deficient weight matrix, next we aim to validate the two proposed strategies, based on natural and artificial rank deficiency of the weight matrix, for 3D line fitting using simulated linear points. The estimated line parameters from the two methods are verified against the true set of parameters used for generating the simulated 3D points. The two approaches are also applied to fit a 3D line for linear features extracted from a real point cloud captured by a Velodyne HDL-32E lidar unit aboard a mobile mapping system. The results from the proposed strategies are compared against those obtained from the TLS approach of Snow and Schaffrin (2016).

**Simulation Results**

The simulated 3D points were generated for multiple line orientations to demonstrate the feasibility of the proposed approach irrespective of line orientation. A total of 25 simulated lines were used. The points along each line were generated using the simulated 3D line parameters and equally spaced along the line. The 3D point coordinates were then contaminated with a noise level of \(2-3\) cm. Figure 3 shows a schematic diagram of the simulated 3D lines. Table 1 (see next page) lists the results obtained for 3D line fitting from the two proposed approaches—natural rank deficiency-based and artificially induced rank deficiency-based—as well as the TLS approach proposed by Snow and Schaffrin (2016). Since the line parameters obtained from the three approaches are exactly the same, they are only listed once in the table. The table includes the total number of points along each line, the major axis of spread \((x, y,\) or \(z)\) of the points, the estimated 3D point \((x_0, y_0, z_0)\) along the line, the line direction vector \((u_x, u_y, u_z),\) and the associated \(a\) posteriori variance factors. The second approach, using an artificially induced rank-deficient weight matrix, requires two of the six line parameters to be fixed. The parameters to be fixed are decided based on the major axis of the line direction, as already discussed in detail. In this regard, the column listing the major axis of spread of the points is used to define the coordinate of the point on the line that is fixed to 0 and the component of the line direction vector that is fixed to 1. Note that Table 1 lists the square root of the \(a\) posteriori variance factors for the first approach; those for each of the two steps involved in the second approach; and those for the TLS approach. The results from both the proposed approaches for line fitting are found to be in accordance with the true values of the parameters used for simulated data. The average difference between the true and estimated point along the line was \((-0.005, 0.023, -0.037)\) m, and that for the line direction vector was on the order of \(1 \times 10^{-5}\) in the \(x, y,\) and \(z\) directions. Furthermore, the proposed approaches are found to be in agreement with the state-of-the-art TLS 3D line fitting approach. Moreover, the \(a\) posteriori variance factors for all the approaches are close to \((3\) cm\(^2\)), which is coherent with the simulated noise level in the 3D points along the features, thus indicating the validity of the estimated 3D line parameters.

![Figure 3. Simulated 3D lines with line identifications.](image-url)

These results indicate equally efficient performance of the two proposed approaches and thus validate the proposition that a naturally occurring rank-deficient weight matrix and an artificially induced weight matrix attain the same solution—the former eliminates noncontributing error terms from the estimation model, whereas the latter eliminates nuisance parameters. For the purpose of 3D line fitting addressed in this case study, we recommend the one-step approach using an LSA model with a naturally occurring rank-deficient weight matrix, since it does not require any prior information regarding approximate line orientation. The performance of the proposed approaches is also found to be on par with the state-of-the-art 3D line fitting strategy. Note, though, that the research presented in this article is not introduced to achieve an improvement over the results of the existing 3D line fitting algorithm. Instead, our key contribution is that the proposed research allows for the formulation of LSA models for generic problems (such as 3D line fitting and a plethora of other problems in the field of geomatics and other engineering fields) without requiring \(a\) priori knowledge of the minimal parameters required to adequately define the model. The current state-of-the-art approaches, including
that of Snow and Schaffrin (2016), rely on a minimal parametrization of 3D lines to obtain accurate line fitting results.

Experimental Results
Three different linear features—two light poles and one stop-sign pole—captured from a Velodyne HDL-32E lidar sensor aboard a mobile mapping system are used to demonstrate the performance of line fitting using real point cloud data. The extracted 3D points and their corresponding imagery are shown in Figure 4. Table 2, which follows the same structure as Table 1, presents the 3D line fitting results obtained for these features using the three approaches. As before, the line parameters are similar from the different approaches, and hence are only stated once. The square root of the a posteriori variance factor is provided for each approach separately. The results indicate an agreement between the two proposed approaches and the TLS approach of Snow and Schaffrin (2016), thus validating the feasibility of the proposed research for conducting accurate 3D line fitting for linear features within real point clouds.

Discussion: Other Applications in Geomatics
As mentioned in the introduction of this article, LSA with a rank-deficient weight matrix is commonly encountered in a vast range of applications, especially in the field of geomatics. This section illustrates the manifestation of such LSA models within a few common applications: plane fitting, mobile lidar system calibration, registration of overlapping 3D point clouds, and single-photo resection. It is worth mentioning that the proposed LSA model is not introduced to outperform existing algorithms for problems in the field of geomatics and other engineering domains. In this regard, the experimental results supporting the applications discussed in this section are not compared to state-of-the-art approaches, since the comparative performance would be the same. Instead, the goal is to emphasize the applicability of the proposed generic LSA model with rank-deficient weight matrix to all such problems without requiring a priori knowledge of the minimal parameters required to adequately define the model. Furthermore, the forthcoming discussion highlights the fact that the rank-deficient weight matrix in the LSA model is indicative of the effective contribution of observations to the estimation process instead of the superficial contribution indicated by the total number of equations, thus facilitating a better understanding of the problem.

3D Plane Fitting
3D plane fitting is a well-addressed problem in various aspects of engineering. One of the mathematical models for solving it uses the standard equation of a plane:

\[ AX + BY + CZ + D = 0 \]

which results in one observation equation per 3D point. Here, \( A, B, C, \) and \( D \) are the plane parameters, three of which are independent. This subsection aims to illustrate that the effective contribution of a 3D point would be one equation even using an alternate LSA model that superficially indicates a contribution of three per point. Moreover, this discussion demonstrates that such an inference can be drawn by assessing the weight matrix of the LSA model. The objective of 3D plane fitting is to identify the plane that minimizes the squared sum of normal distances from the given 3D points to the best-fitting plane. Based on the schematic illustration shown in Figure 5,
let us assume that the plane is defined by an arbitrary point $\alpha$ on it, whose coordinates are given as $(x_0, y_0, z_0)$, and a unit direction vector $(w_x, w_y, w_z)$ normal to the plane. For any observed point $b$ with coordinates $(x_i, y_i, z_i)$, its projection on the plane is denoted as $b_n(x_n, y_n, z_n)$, and the normal distance vector of the point from the plane is denoted as $(d_x, d_y, d_z)$.

The noise-free 3D point coordinates are denoted by $(x_i, y_i, z_i)$:

$$\begin{bmatrix}
  x_i \\
  y_i \\
  z_i 
\end{bmatrix} = \begin{bmatrix}
  x_i - e_{x_i} \\
  y_i - e_{y_i} \\
  z_i - e_{z_i} 
\end{bmatrix}$$

where $(e_{x_i}, e_{y_i}, e_{z_i})$ refers to the noise in the 3D point coordinates. The distance vector from point $b$ to the plane can be written as

$$\begin{bmatrix}
  d_x \\
  d_y \\
  d_z 
\end{bmatrix} = \begin{bmatrix}
  x_i \\
  y_i \\
  z_i 
\end{bmatrix} - \begin{bmatrix}
  x_0 \\
  y_0 \\
  z_0 
\end{bmatrix}$$

Table 2. 3D line fitting results for real point cloud data from three approaches.

<table>
<thead>
<tr>
<th>Line</th>
<th>Number of Points</th>
<th>Major Axis of</th>
<th>Point $(x_0, y_0, z_0)$ on Line (m)</th>
<th>Line Direction $(u_x, u_y, u_z)$</th>
<th>Approach 1: $\delta(x)$ (cm)</th>
<th>Approach 2: $\delta_{\text{point}}, \delta_{\text{dir}}$ (cm)</th>
<th>Approach 3: $\delta_z$ (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>27 114</td>
<td>Z</td>
<td>(500 400.085, 4 479 885.347, 0.000)</td>
<td>$(-0.004, -0.000, 1.000)$</td>
<td>3.1</td>
<td>3.1, 3.1</td>
<td>3.1</td>
</tr>
<tr>
<td>L2</td>
<td>21 419</td>
<td>Z</td>
<td>(500 412.932, 4 479 903.172, 0.000)</td>
<td>$(0.012, -0.008, 1.000)$</td>
<td>3.4</td>
<td>3.4, 3.4</td>
<td>3.4</td>
</tr>
<tr>
<td>L3</td>
<td>10 015</td>
<td>Z</td>
<td>(500 403.468, 4 479 904.699, 0.000)</td>
<td>$(-0.009, 0.000, 1.000)$</td>
<td>2.3</td>
<td>2.3, 2.3</td>
<td>2.3</td>
</tr>
</tbody>
</table>

*a* Natural rank deficiency.

*b* Artificially induced rank deficiency.

*c* Total least squares (Snow and Schaffrin 2016).

**Figure 4.** 3D linear features (colored by height) extracted from real point cloud data.

**Figure 5.** Schematic illustration of 3D plane fitting by minimizing the normal distances between the given points and fitted 3D plane.
Expanding Equation 47 results in a set of three equations:

\[
\begin{align*}
d_x &= \bar{x} w_x^2 + \bar{y} w_x w_y + \bar{z} w_x w_z - x_0 \\
d_y &= \bar{x} w_y w_x + \bar{y} w_y^2 + \bar{z} w_y w_z - y_0 \\
d_z &= \bar{x} w_z w_x + \bar{y} w_z w_y + \bar{z} w_z^2 - z_0
\end{align*}
\] (48a)

(48b)

(48c)

Linearization of the mathematical model in Equation 48 results in the Gauss–Helmert model with matrix \( B \), composed of the partial derivatives of the function with respect to the observations:

\[
B = \begin{bmatrix}
\frac{\partial d_x}{\partial x}, & \frac{\partial d_x}{\partial y}, & \frac{\partial d_x}{\partial z} \\
\frac{\partial d_y}{\partial x}, & \frac{\partial d_y}{\partial y}, & \frac{\partial d_y}{\partial z} \\
\frac{\partial d_z}{\partial x}, & \frac{\partial d_z}{\partial y}, & \frac{\partial d_z}{\partial z}
\end{bmatrix} = \begin{bmatrix}
w_x^2 & w_y w_x & w_z w_x \\
w_x w_y & w_y^2 & w_z w_y \\
w_x w_z & w_y w_z & w_z^2
\end{bmatrix}
\] (49)

Although this mathematical model seems to indicate a contribution of three observation equations from each point, an in-depth assessment of the resultant LSA model and the rank-deficient weight matrix indicates an effective contribution of only one observation equation per 3D point. Intuitively, this contribution of a single observation equation corresponds to the random error in the direction normal to the plane.

It can be seen that the matrix \( B \) is rank-deficient with rank 1, since each row can be expressed as a scalar multiple of either of the remaining two rows. Now, the variance-covariance matrix of the combined error term \( \Sigma_e \), as derived using the law of error propagation, is given by the following equation, where \( \sigma_{sys} \) denotes the standard deviation of the 3D point coordinates:

\[
\Sigma_e = \sigma_{sys}^2 B B^T
\] (50)

Assuming identical weights for all the 3D points (that is, \( P = I \)), the term \( \Sigma_e \) can be further simplified:

\[
\Sigma_e = \sigma_{sys}^2 B^T B
\] (51)

To assess the effective contribution of a set of observations, an in-depth analysis of the weight matrix associated with the combined error term is conducted. The weight matrix can be obtained as the inverse of the variance-covariance matrix \( \Sigma_e \), the desired weight matrix of the combined error term is obtained by computing the Moore–Penrose pseudo-inverse of \( \Sigma_e \). This is achieved by conducting an eigenvalue decomposition of the matrix \( B \). The eigenvectors of a symmetric matrix form a set of orthonormal basis vectors (Osnaga 2005), the remaining two eigenvectors will correspond to the two directions \( u \) and \( v \) that lie along the plane, as shown in Figure 5. The derived eigenvectors of the matrix \( B \) indicate that the matrix \( V \) of eigenvectors actually represents the rotation matrix from the local 3D plane coordinate system to the mapping frame coordinate system (as illustrated in Figure 5), or \( R_{uvw} \).

Analyzing the combined error term \( \Sigma_e \) (or \( B \Sigma_e \)) by substituting the results from the eigenvalue decomposition of \( B \) results in

\[
\begin{bmatrix}
e_x \\
e_y \\
e_z
\end{bmatrix} = \begin{bmatrix}
R_{uvw} & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
e_x \\
e_y \\
e_z
\end{bmatrix}
\] (52)

This simplification indicates that the rank-deficient weight matrix indeed nullifies the error components \( e_x \) and \( e_y \) in the orthogonal directions that lie along the plane, while only considering the error \( e_z \) in the normal direction to the plane. This in turn indicates that the effective contribution from each set of observations from a 3D point is only one equation, as indicated by the rank of the matrix \( B \) and interpreted using the eigenvector matrix of \( B \). The three plane parameters being estimated in this LSA model can be interpreted as the point of intersection of the plane with the \( x \), \( y \), or \( z \)-axis (one parameter) and the direction vector normal to the plane (two parameters).

The proposed approach is validated using a real point cloud data set. Three different planar features—two building facades and one ground patch—captured by a Velodyne HDL-32E lidar sensor aboard a mobile mapping system are used to demonstrate the performance of plane fitting using real point cloud data. The extracted planar patches and their corresponding imagery are shown in Figure 6. Table 3 presents the plane fitting results obtained for these features using the proposed LSA strategy with rank-deficient weight matrix. The table includes the total number of points along each plane, major axes of spread (\( xy \), \( yz \), or \( xz \)) of the points, estimated 3D point (\( x_0 \), \( y_0 \), \( z_0 \)) along the plane, normal direction vector (\( w_x \), \( w_y \), \( w_z \)), and associated \( a \) posteriori variance factor. Here the column listing the type of planar feature facilitates an understanding of the two coordinates of the point on the plane that are fixed to 0 and the component of the normal direction vector that is fixed to 1. The \( a \) posteriori variance factor is observed to lie within the noise level of the 3D point cloud, thus validating the feasibility of the proposed research for accurate plane fitting.

**Feature-Based Mobile Lidar System Calibration**

The field of geomatics has expanded in recent years to explore the use of mobile mapping systems equipped with lidar and global navigation satellite system/inertial navigation system (GNSS/INS) units for various applications, such as transportation corridor mapping (Caltagirone et al. 2019; Ravi et al. 2020), precision agriculture (Koenig et al. 2015; Ravi, Lin, Shamseldin et al. 2018; Ravi et al. 2019), infrastructure monitoring (Soilán et al. 2019; Aldosari et al. 2020; Al-Rawabdeh et al. 2020), shoreline monitoring (Flener et al. 2013; Lin et al. 2019), and archaeological mapping (Inomata et al. 2017; Lin et al. 2019). However, one of the most crucial steps in deriving accurate 3D point clouds is calibrating the system, which entails the intrinsic parameters of onboard sensors as well as the extrinsic or mounting parameters relating the different sensors to the GNSS/INS unit. This discussion emphasizes rank-deficient weight matrices within LSA models that are artificially induced to eliminate nuisance parameters from the model and obtain an accurate estimate of the desired system calibration parameters.

The schematic illustration in Figure 7a depicts a misalignment between 3D point clouds from different tracks, in the case of an uncalibrated system. On the other hand, Figure 7b shows the expected impact of system calibration toward achieving good alignment of 3D point clouds. A mobile lidar system is calibrated by minimizing the misalignment between 3D point clouds captured from different tracks or sensors. To this end, this subsection includes a discussion of the calibration approach and its formulation involving an artificially induced rank-deficient weight matrix.

Several calibration strategies have been proposed in recent years, most of which rely on specially designed features or those naturally occurring in the surrounding environment.
Table 3. Plane fitting results for real point-cloud data.

<table>
<thead>
<tr>
<th>Plane</th>
<th>Number of Points</th>
<th>Type of Planar Feature</th>
<th>Point ((x_0, y_0, z_0)) on Plane (m)</th>
<th>Normal Vector Direction ((w_x, w_y, w_z))</th>
<th>(\hat{\sigma}_0) (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PW1</td>
<td>10 116</td>
<td>yz (building facade)</td>
<td>((48 8075.846, 0.000, 0.000))</td>
<td>((1.000, -0.006, -0.007))</td>
<td>1.3</td>
</tr>
<tr>
<td>PW2</td>
<td>9963</td>
<td>xz (building facade)</td>
<td>((0.000, 4 482 366.818, 0.000))</td>
<td>((0.066, 1.000, 0.009))</td>
<td>1.3</td>
</tr>
<tr>
<td>PG1</td>
<td>10 187</td>
<td>xy (ground patch)</td>
<td>((0.000, 0.000, -81 942.232 279 262))</td>
<td>((-0.011, -0.017, 1.000))</td>
<td>1.2</td>
</tr>
</tbody>
</table>

Figure 6. 3D planar features (black) extracted from real point-cloud (colored by height).

Figure 7. Schematic illustration of alignment between point clouds from different tracks: (a) bad alignment for uncalibrated mobile mapping system; (b) good alignment for accurately calibrated mobile mapping system.
The strategy proposed by Ravi, Lin, Elbahnasawy et al. (2018) for estimating mounting parameters (lever arm and boresight angles) relies on minimizing the discrepancy between corresponding points or linear and planar features captured from different tracks and/or sensors. This is achieved by minimizing the discrepancy in 3D coordinates between each conjugate 3D point pair as well as the nonconjugate point pairs formed between corresponding features, resulting in three observation equations per point pair:

\[
\begin{bmatrix}
X \\
Y \\
Z
\end{bmatrix}_{\text{track1}} - \begin{bmatrix}
X \\
Y \\
Z
\end{bmatrix}_{\text{track2}} = \begin{bmatrix}
e_x \\
e_y \\
e_z
\end{bmatrix}
\] (53a)

Linear/planar features:

\[
\begin{bmatrix}
X \\
Y \\
Z
\end{bmatrix}_{\text{track1}} - \begin{bmatrix}
X \\
Y \\
Z
\end{bmatrix}_{\text{track2}} = \begin{bmatrix}
d_x \\
d_y \\
d_z
\end{bmatrix} + \begin{bmatrix}
e_x \\
e_y \\
e_z
\end{bmatrix}
\] (53b)

Here, \((e_x, e_y, e_z)\) denotes the random error in the discrepancy between the 3D point pair. The 3D coordinates \((X, Y, Z)\), for a point are expressed as a function of the system measurements in the corresponding track and the intrinsic and extrinsic system calibration parameters. Note that for linear and planar features, each pair is formed between nonconjugate points along corresponding features, resulting in a nonrandom component of discrepancy lying along the linear or planar feature, which is denoted by \((d_x, d_y, d_z)\) in Equation 53b and shown as \(\Delta E\) in Figure 8. So a modified weight matrix is introduced in the LSA model in such cases to nullify the vector \((d_x, d_y, d_z)\) or the components of the discrepancy between the constituents of a point pair that are not normal to the corresponding linear or planar feature. The modified weight matrix is designed in the local coordinate frame \((uvw)\) defined according to the linear or planar feature, similar to the prior discussion for 3D line and plane fitting. The resultant modified weight matrix will be rank-deficient with rank 2 for linear features and rank 1 for planar features, resulting in an LSA model with an artificially induced rank-deficient weight matrix. Note that such a weight modification is not required for conjugate 3D point pairs.

An alternate approach to calibrating a mobile lidar system that we have proposed (Ravi and Habib 2020) relies on minimizing the discrepancy between corresponding thin profiles extracted automatically from different tracks. That approach is based on nonconjugate 3D point pairs formed along corresponding profiles, resulting in a discrepancy expressed similarly to Equation 53b. In this case, the nonrandom component of the discrepancy lies in the direction across the profile length, as shown in Figure 9. This in turn mandates minimizing the discrepancy between nonconjugate point pairs along corresponding profiles only along the profile length and in the vertical directions. This is achieved by introducing a rank-deficient modified weight matrix (rank 2) in the local coordinate frame of the profile, similar to the approach adopted for linear features, where the weight in the normal direction to the profile \(P\) was set to 0. Therefore, the effective contribution of a 3D point pair to redundancy is reduced to only two equations instead of three.

The proposed LSA model with rank-deficient weight matrix for mobile lidar system calibration is validated by conducting a feature-based calibration for a Velodyne HDL-32E sensor aboard a mobile mapping system. Nine planar features were used to conduct the calibration, as shown in Figure 10. The nominal and calibrated mounting parameters—lever arm \((\Delta X, \Delta Y, \Delta Z)\) and boresight angles \((\Delta \alpha, \Delta \phi, \Delta \kappa)\)—are provided in...
Table 4. The square root of the a posteriori variance factor is 1.8 cm, which is in agreement with the expected accuracy of 2–3 cm for the captured point clouds based on the onboard sensor specifications. These results prove that the proposed LSA model with rank-deficient weight matrix can accurately calibrate mobile lidar systems.

Registration of Overlapping 3D Point Clouds
Another application frequently encountered while dealing with 3D point clouds in the field of geomatics is the registration of overlapping point clouds to ensure good alignment (or minimal discrepancy) between them. This problem is most often encountered when working with 3D point-clouds captured by terrestrial laser scanners, which result in point clouds referenced to a different local coordinate system for each station set up to capture different segments of the scene of interest, as shown schematically in Figure 11a. Reconstructing the complete scene using the point clouds captured from the different stations requires the registration of these point clouds to a common reference frame, as shown in Figure 11b.

The problem of 3D registration has been addressed in prior research (Jaw and Chuang 2008; Al-Durgham and Habib 2014; Yang et al. 2016; Frokop et al. 2020); the purpose of the discussion in this subsection is to demonstrate the use of artificially induced rank-deficient weight matrices in LSA models for this application and to interpret this use in terms of its impact on the effective contribution of 3D points to the solution. Minimizing the discrepancy between overlapping point clouds is the basis for 3D point-cloud registration as well as mobile lidar system calibration, already discussed, meaning we can use the same mathematical model as in Equation 53a.

In the case of registration, however, the 3D point coordinates are a function of the transformation parameters (translation, rotation, and scaling factor) estimated during registration, and the discrepancy between the transformed coordinates is minimized for a point captured in two different scans. Similar to mobile system calibration, commonly used registration primitives include points as well as 3D linear and planar features, which are used to form nonconjugate 3D point pairs to quantify the discrepancy that needs to be minimized. Again, the nonconjugate point pairings would result in a nonrandom discrepancy component and thus the revised mathematical model given in Equation 53b. This mandates an artificially modified rank-deficient weight matrix to nullify the discrepancy components that are not normal to the corresponding feature. An assessment of the resultant LSA model and its rank-deficient weight matrix for feature-based registration (similar to that conducted for system calibration) implies a reduced effective contribution of 3D point pairs (two equations for linear features and one for planar) to the redundancy of transformation parameters estimation.

The proposed LSA model for feature-based registration was used to conduct a fine registration between 3D point

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Type</th>
<th>ΔX (m)</th>
<th>ΔY (m)</th>
<th>ΔZ (m)</th>
<th>Δω (°)</th>
<th>Δϕ (°)</th>
<th>Δκ (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal</td>
<td></td>
<td>−1.05</td>
<td>−0.45</td>
<td>−0.32</td>
<td>180</td>
<td>−15</td>
<td>−0</td>
</tr>
<tr>
<td>Calibrated</td>
<td></td>
<td>−1.0852</td>
<td>−0.4522</td>
<td>−0.32</td>
<td>180.0874</td>
<td>−18.8267</td>
<td>−0.3232</td>
</tr>
</tbody>
</table>

Figure 10. Calibration features used for Velodyne HDL-32E lidar sensor aboard a mobile mapping system.

Figure 11. 3D registration: (a) scanning stations (top view) and (b) sample 3D point clouds captured from different stations before and after registration (perspective view).
clouds obtained from two static terrestrial laser scanners (a FARO Focus 3D X330 and Trimble TX8b). A total of six scans were registered using planar, linear, and cylindrical features, shown in blue, red, and green, respectively, in Figure 12. The registration results improved the root-mean-square error of the distances of the points from the corresponding best-fitting feature from 1.4 cm to 0.4 cm. Further, the square root of the $a$ posteriori variance factor was 0.6 cm, which indicates the feasibility of the proposed LSA model with rank-deficient weight matrix for 3D point-cloud registration.

**Single-Photo Resection**

The objective of single-photo resection is to determine the exterior orientation parameters (EOPs) for a given image relative to the ground coordinate system. This is achieved using control points (Habib and Mazaheri 2015) or control lines (Habib et al. 2003), illustrated schematically in Figure 13a and 13b, respectively. Single-photo resection using control points relies on an LSA model whose objective function is based on enforcing the collinearity of the perspective center, an object-space (or ground) point, and the corresponding image-space point. On the other hand, using control lines, Habib et al. proposed two different approaches using different representation schemes for control lines: a sequence of 2D and 3D points along the linear feature in image and object spaces, respectively, and polylines (a sequence of straight-line segments) to represent features. Their first representation scheme used exact point-to-point correspondences between 2D and 3D points along the linear feature to apply LSA based on collinearity equations. The traditional collinearity equations used to solve the problem of LSA using control points are given by

$$
x = x_0 - \frac{N_x}{D} \text{dist}_x + \sigma_x; \quad y = y_0 - \frac{N_y}{D} \text{dist}_y + \sigma_y; \quad \text{where } \begin{bmatrix} \sigma_x \\ \sigma_y \end{bmatrix} = (0, \sigma_{w}P_v) \tag{54}
$$

$$
N_x = r_{11}(X - X_0) + r_{21}(Y - Y_0) + r_{31}(Z - Z_0) \tag{55a}
$$

$$
N_y = r_{12}(X - X_0) + r_{22}(Y - Y_0) + r_{32}(Z - Z_0) \tag{55b}
$$

$$
D = r_{13}(X - X_0) + r_{23}(Y - Y_0) + r_{33}(Z - Z_0) \tag{55c}
$$

where $(x, y)$ and $(X, Y, Z)$ denote the 2D image coordinates and 3D ground coordinates of the control point; $(x_0, y_0)$ denotes the principal point coordinates; $c$ denotes the principal distance; $(\text{dist}_x, \text{dist}_y)$ denotes the distortion in image coordinates; $(X_0, Y_0, Z_0)$ denotes the position of the image with respect to the ground coordinate frame; and $r_{ij}$ denotes the $(i, j)^{th}$ element of the rotation matrix relating the two frames. Hence, each conjugate point pair results in two observation equations contributing to the solution. In the polyline-based representation scheme, single-photo resection is carried out using an objective function that enforces the coplanarity of the perspective center, image-space line, and object-space line—that is, the coplanarity of the vectors $v_a$, $v_A$, and $v_p$ as shown in Figure 13b.

While the model based on collinearity equations relies on conjugate ground and image point pairs, establishing such a correspondence between the set of points along linear features in two different spaces (image space and object space) is challenging. In this discussion, we illustrate the introduction of an artificially modified weight matrix to adapt the collinearity equations-based approach to suggest a point-based incorporation of control lines for single-photo resection with no requirement for exact point-to-point correspondences. So, an alternate pairing scheme using nonconjugate points along corresponding linear features is used. Taking Figure 14 as reference, the ground points $A_i$ lying along the object-space
line are paired with nonconjugate image points \( a_i \) lying along the corresponding image-space line. Such a pairing scheme results in a nonrandom discrepancy between the projected ground and image point pairs that lies along the image line direction, as shown by the vector \( d_i \) in Figure 14. Hence, the collinearity equations are modified as follows to include the nonrandom discrepancy \( (d_i, d_j) \) along the image-space line:

\[
x = x - c \frac{N}{D} \text{dist}_i + d_i + e_i \quad y = y - c \frac{N}{D} \text{dist}_j + d_j + e_j \quad \text{where} \quad \epsilon_i = \epsilon_j = 0.5 \sigma_P^2 \quad (56)
\]

Based on the 2D line parameters defining the image line, a modified weight matrix can be generated in the local coordinate frame \( uvw \) of the 2D line (as shown in Figure 14) to eliminate the nonrandom discrepancy lying along the image line direction. This modified weight matrix would be rank-deficient with rank 1. The resultant rank deficiency thus indicates an effective contribution of one equation for a 2D point measured along a control line instead of the regular contribution of two equations from a control point.

Single-photo resection based on control lines with nonconjugate point pairings is used to estimate the EOPs of an image in Figure 15. The coordinates of the centers of the checkerboard targets are known in the ground reference frame. The checkerboards lying along the same edge of the calibration cage are used to define the control lines for single-photo resection. For instance, the close-up view in Figure 15 shows checkerboard centers labeled C1–C5 constitute the control line L1. To test the proposed approach, the image coordinates measured for the checkerboards along a control line are paired with nonconjugate ground coordinates of another checkerboard target lying along the same control line. For instance, the image coordinates of C1 are paired with ground coordinates of C2, since they lie along the same control line, and so on. The EOPs for the same image were also estimated using the traditional point-based approach. The estimated EOPs—position \( (X, Y, Z) \) and orientation \( (\omega, \phi, \kappa) \)—from the two approaches are provided in Table 5. The square root of the a posteriori variance factor is 1.41 and 1.32 pixels for the traditional and control-line approaches, respectively. The similarity of the results obtained from the two approaches validates the proposed LSA model for single-photo resection based on control lines.

**Conclusions**

This article started by discussing LSA models with a rank-deficient weight matrix and proposed a solution to such models by building on the solutions to existing common LSA models (Gauss–Markov and Gauss–Helmert models). Furthermore, the prominence of this problem and its solution were explained in more detail using a case study pertaining to 3D line fitting. This case study proposed two different approaches to deriving an accurate solution to the 3D line fitting problem—one using an LSA model with naturally rank-deficient weight matrix and another, two-step approach based on an artificially induced rank-deficient weight matrix. The study also provided critical insight into the mathematical and physical interpretation of the sources of rank deficiency. The feasibility of the resultant solution was verified by applying the proposed strategy to estimate line parameters in simulated data. The results showed agreement of the estimated parameters with the true parameters used for simulation.

Finally, the manifestation of naturally occurring and artificially induced rank-deficient weight matrices was illustrated using various geomatics-related applications: 3D plane fitting, mobile lidar system calibration, 3D point-cloud registration, and single-photo resection. The discussion of these applications aimed to emphasize the need for assessment of the weight matrix involved in LSA models to gain a thorough understanding of the effective contribution of observations to the solution of the problem at hand. Furthermore, the research proposed in this article facilitates the option of artificially introducing a rank-deficient weight matrix in LSA models to adapt existing traditional techniques in geomatics that rely mainly on points to also incorporate higher-level primitives, such as linear and planar features, according to the application of interest. Note that the stated inference is not limited to the aforementioned applications, and can be easily applied to a multitude of other problems related to lidar, photogrammetry, and remote sensing. The work presented in this article provides useful tools and insight to advance research in various fields of engineering, which was not possible earlier due to lack of an in-depth understanding of and solutions to such LSA models.
References


Appendix: Least Squares Adjustment with a Rank-Deficient Weight Matrix

Let us start with a Gauss–Markov LSA model involving a rank-deficient weight matrix:

\[ y = Ax + e; \quad e \sim (0, \sigma_0^2 P') \] (A.1)

Here \( y \) denotes the \( n \times 1 \) vector of observations with an associated \( n \times n \) rank-deficient weight matrix denoted by \( P' \). \( A \) is the \( n \times m \) design matrix, \( x \) is the \( m \times 1 \) vector of unknown parameters (with \( n > m \)), and \( \sigma_0^2 \) is the \textit{a priori} variance factor. The corresponding error term \( e \) is given by

\[ e = y - Ax \] (A.2)

The objective of LSA is to minimize \( e^T P'e \), which leads to the following:

\[ \phi(x) = e^T P'e = (y - Ax)^T P' (y - Ax) = \min_x \] (A.3)

\[ = y^T P'y - 2x^T A^T P'y + x^T A^T P'Ax \]

We differentiate the objective function with respect to the unknown parameters \( x \) and equate it to 0:

\[ \frac{\partial \phi}{\partial x} = -2A^T P'y + 2A^T P'A \hat{x} = 0 \] (A.4)

This gives us the least squares estimate of the unknowns:

\[ \hat{x} = (A^T P'A)^{-1} A^T P'y \] (A.5)

Using the law of error propagation, the variance-covariance matrix (or dispersion matrix) of the estimated unknown parameters is given by

\[ \Sigma(x) = \sigma_0^2 ((A^T P'A)^{-1} A^T P'P'(P'A(A^T P'A)^{-1})) \] (A.6a)

which—since, for a Moore–Penrose pseudo-inverse, \( P'P' = P \) (Penrose 1955)—can be further simplified to

\[ \Sigma(x) = \sigma_0^2 (A^T P'A)^{-1} \] (A.6b)

The \textit{a posteriori} variance factor \( \hat{\sigma}^2 \) is obtained by deriving the expected value of the sum of squares of the weighted predicted residuals, as given by

\[ E(\hat{e}^T P' \hat{e}) = E \left\{ (y - A \hat{x})^T P'(y - A \hat{x}) \right\} \] (A.7a)

by substituting the expression for \( \hat{x} \), considering that \( (I - AN^{-1} A^T P') \) is an idempotent matrix, where \( N = A^T PA \). Given that the trace of a scalar equals the scalar, and that the trace operation is commutative, Equation A.7b can be manipulated to get

\[ E(\hat{e}^T P' \hat{e}) = E \left\{ \text{tr} (P'y^T) - \text{tr} (P'AN^{-1} A^T P'y^T) \right\} \] (A.7c)

which can be further simplified, using the properties of trace operations, to

\[ E(\hat{e}^T P' \hat{e}) = \text{tr} (P'(I_n - AN^{-1} A^T P') E \left\{ y y^T \right\} ) \] (A.7d)

The term \( E \left\{ yy^T \right\} \) can be derived from the variance-covariance matrix of the observations vector \( \Sigma(y) \):

\[ \Sigma(y) = \sigma_0^2 P' = E((y - Ax)(y - Ax)^T) \] (A.8a)

\[ \Rightarrow \sigma_0^2 P' = E(yy^T) - Axx^TA^T \] (A.8b)

\[ \Rightarrow E(yy^T) = \sigma_0^2 P' + Axx^TA^T \] (A.8c)

Substituting this into Equation A.7d and simplifying yields

\[ E(\hat{e}^T P' \hat{e}) = \sigma_0^2 \text{tr} (PP') - \sigma_0^2 \text{tr} \left( (A^T P'A)^{-1} A^T P' \right) \] (A.9a)

which can be simplified further to give

\[ E(\hat{e}^T P' \hat{e}) = \sigma_0^2 \text{rank}(P') - \sigma_0^2 \frac{m}{\text{rank}(P') - m} \] (A.9b)

using the fact that \( \text{tr}(PP') = \text{rank}(PP') = \text{rank}(P) \), since it is an idempotent matrix. Finally, the estimated \textit{a posteriori} variance factor is given by

\[ \hat{\sigma}^2 = \frac{\hat{e}^T P' \hat{e}}{\text{rank}(P') - m} \] (A.10)

indicating that the redundancy for such a rank-deficient LSA model is defined as the difference between the rank of the weight matrix and the number of unknown parameters \( m \). Thus, the term \( \text{rank}(P') - m \) in the denominator of Equation A.10 indicates the contribution of equations to the redundancy.
IN-PRESS ARTICLES

Abdullah Kayı, Bülent Bayram, and Dursun Zafer Şeker. The Analysis on the Annual Change of Digital Aerial Camera’s IMUs Boresight Misalignment.


Binbin Li, Huan Xie, Shijie Liu, Xiaohua Tong, Hong Tang, and Xu Wang. A Method of Extracting High-Accuracy Elevation Control Points from ICESat-2 Altimetry Data.

Forrest Corcoran and Christopher E. Parrish. Diffuse Attenuation Coefficient (Kd) from ICESat-2 ATLAS Spaceborne Lidar Using Random-Forest Regression.

S. Boukir, L. Guo, and N. Chehata. Improving Remote Sensing Multiple Classification by Data and Ensemble Selection.


Bo Yu, Fang Chen, Ying Dong, Lei Wang, Ning Wang, and Aqiang Yang. MSegnet, a Practical Network for Building Detection from High Spatial Resolution Images.

Feizhou Zhang, Xufang Liu, Yun Xing, Zihan Zhang, Siyuan Liu, and Lei Yan. Estimation of Rock Characteristics Based on Polarization Spectra: Surface Roughness, Composition, and Density.


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

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

Steven Martinez Vargas, Claudio Delrieux, Katy L. Blanco, and Alejandro Vitale. Dense Bathymetry in Turbid Coastal Zones Using Airborne Hyperspectral Images.


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

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

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

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


Xuzhe Duan, Qingwu Hu, Pengcheng Zhao, and Shaohua Wang. A low-cost and portable indoor 3D mapping approach using biaxial line laser scanners and a one-dimension laser rangefinder integrated with MEMS.
Spectral Reflectance Estimation of UAS Multispectral Imagery Using Satellite Cross-Calibration Method

Saket Gowravaram, Haiyang Chao, Andrew Molthan, Tiebiao Zhao, Pengzhi Tian, Harold Flanagan, Lori Schultz, and Jordan Bell

Abstract
This paper introduces a satellite-based cross-calibration (SCC) method for spectral reflectance estimation of unmanned aircraft system (UAS) multispectral imagery. The SCC method provides a low-cost and feasible solution to convert high-resolution UAS images in digital numbers (DN) to reflectance when satellite data is available. The proposed method is evaluated using a multispectral data set, including orthorectified KHawk UAS DN imagery and Landsat 8 Operational Land Imager Level-2 surface reflectance (SR) data over a forest/grassland area. The estimated UAS reflectance images are compared with the National Ecological Observatory Network’s imaging spectrometer (NIS) SR data for validation. The UAS reflectance showed high similarities with the NIS data for the near-infrared and red bands with Pearson’s r values being 97 and 95.74, and root-mean-square errors being 0.0239 and 0.0096 over a 32-subplot hayfield.

Introduction
Unmanned aircraft systems (UAS) have been widely used for many multispectral remote sensing applications, including tornado damage track identification (Wagner et al. 2019; Gowravaram et al. 2018), agriculture mapping (Torres-Rua et al. 2015; Niu et al. 2019; Zhao et al. 2015), and forest fire monitoring (Ononye et al. 2007; Merino et al. 2012; Fraser et al. 2017). UAS aerial images can facilitate real-time observations for regions of interest at high spatiotemporal resolutions (e.g., submeter spatial resolution and hourly temporal resolution), which can enable small-scale feature detection that may or may not be visible in satellite or aircraft imagery. They can also provide data under cloudy conditions when satellite and aircraft imagery are obstructed. However, one of the biggest challenges for UAS based multispectral remote sensing is the retrieval of reflectance from raw orthorectified UAS images in digital numbers (DN).

In recent years, there have been notable contributions towards ground-based radiometric calibration for UAS multispectral imagery (Stark et al. 2016; Edwards et al. 2019). The most popular methods use either commercial or customized ground reflectance target boards as a reference before, during, and after the UAS flight to identify the relationship between DN and reflectance for each spectral band. A semi-automatic model-based method was proposed to convert raw UAS images to reflectance using a white barium sulfate panel and a Halon board (Zaman et al. 2014). Similarly, a customized gray gradient Masonite panel with nine different levels was used for UAS image radiometric calibration using a simplified empirical line method (Wang and Myint 2015). Their results showed that the relationship between UAS imagery in DN and reflectance is nonlinear for many commercial off-the-shelf cameras. Recently, a subband empirical line radiometric calibration method was proposed using a six-band miniature multiple camera array by Tetracam (Deng et al. 2018), which used standard diffuse panels as reference boards. The above methods have been widely used in research communities for reflectance estimation of UAS images. However, they are dependent on the accuracy and the Lambertian properties of the reference boards. To avoid this issue, most UAS groups use spectroradiometers before every survey mission to accurately measure the reflectance of the reference boards. These instruments can be expensive and infeasible for many UAS end users who are constrained by a budget. Also, many existing UAS data sets have not been converted to spectral reflectance due to a lack of ground spectroscopy measurements.

Satellites have been the primary source for large-scale multispectral remote sensing applications, including burn severity mapping, earthquake damage assessment, and landslide extent mapping (Stow et al. 2007; Voigt et al. 2007). Free and open satellite data can be obtained from the National Aeronautics and Space Administration’s (NASA) Landsat, Terra, and Aqua satellites, and the European Space Agency’s Sentinel. However, these satellites provide imagery at lower spatial resolutions (10 m or lower) with relatively slow revisit times (1-16 days).

While satellite imagery possesses reliable radiometric accuracy, UAS can acquire images at desirable spatiotemporal resolutions, which has led to efforts towards understanding the relationship between UAS and satellite remote sensing data. High-resolution aerial observations from UAS were used to increase the spatial resolution of satellite data for precision agriculture (Hassan-Esfahani et al. 2017). Similarly, the variations in the normalized difference vegetation index (NDVI) across different pixel scales were studied among canopy and noncanopy vegetation using various sensing payloads, including UAS cameras, Landsat 8 (L8) Operational Land Imager (OLI), and the Moderate Resolution Imaging Spectroradiometer (Wang et al. 2017). Airborne data from manned aircraft has also been explored for multispectral remote sensing.
(Joan-Cristian et al. 2019). Detailed statistics and cost analysis between UAS, manned aircraft, and satellite remote sensing platforms were provided for precision viticulture applications (Mateo et al. 2015) and a satellite-based local bidirectional reflectance distribution function (BRDF) correction method was introduced for radiometric correction of digital aerial images using data from Landsat 5 and 7 (Tuominen and Pekkarinen 2004). While satellite, manned aircraft, and UAS all have their advantages and disadvantages, it is important to understand the radiometric relationships between different remote sensing data from various sources for future integrated and improved earth observations.

For accurate integration and comparison of multisource remote sensing data, UAS and satellite reflectance imagery acquired at the same time are preferred to minimize radiometric differences. However, this can be difficult due to other constraints. In such cases, satellite imagery acquired a few days before or after the UAS campaign can be considered for radiometric studies and comparison. For example, UAS and RapidEye images three days apart were used for precision viticulture applications (Mateo et al. 2015), and L8 images acquired two days after a UAS mission were compared to the UAS reflectance and NDVI values for vegetation monitoring (Berra et al. 2017). Similarly, an adaptive classification approach for precision agriculture monitoring was developed where drone and L8 images acquired seven days apart were used (Murugan et al. 2017).

This paper focuses on the development of a low-cost cross-calibration method for estimation of UAS spectral reflectance at high spatial resolution using satellite reflectance data of the same area which can be beneficial to UAS operators and research groups who want to: (1) collect new UAS data but do not possess accurate spectroradiometers and ground target boards, (2) calibrate existing UAS data collected without a ground reflectance reference, and (3) study the radiometric relationships between multi-scale remote sensing data from satellite, manned aircraft, and UAS for enhanced Earth observations. The proposed method is demonstrated using orthorectified KHawk UAS DN imagery and L8 OLI Level-2 (L2) surface reflectance (SR) data of a forest/grassland area in Kansas. The cross-calibration functions between these two data are first identified for each spectral band and then used to convert high-resolution UAS DN images to reflectance. The proposed satellite-based cross-calibration (SCC) method is finally validated by comparing the estimated UAS reflectance images with National Ecological Observatory Network (NEON) Image spectrometer (NIS) atmospherically corrected SR data.

### UAS and Satellite Remote Sensing Data

This section provides descriptions of the study area and the remote sensing data used to formulate and validate the proposed method.

#### Study Area

An area of 0.5 square kilometers at the University of Kansas Field Station is selected as the study area (Figure 1). It is located at 39.054°–95.190°, elevation 331 m in the deciduous forest and tall grass prairie ecotone of northeastern Kansas (KBS 2019). Figure 1 shows the study area and the corresponding orthorectified UAS image. Note that this area contains a 32-subplot hayfield (labeled in Figure 1) close to the northwest boundary, which is maintained annually by the Kansas Biological Survey (KBS). Each subplot (10×10 m) is given the same treatment (seeded, fertilized, or hayed) and is maintained to have the same vegetation (Foster et al. 2010). This hayfield is well suited for validation of the proposed SCC method. This area is also one of the 81 NEON field sites, a 30-year ecological monitoring project, and one of the most extensive initiatives of the National Science Foundation (NSF) (NEON 2019).

#### KHawk UAS Data

KHawk 2 UAS is a low-cost multispectral remote sensing platform developed by the Cooperative Unmanned Systems Lab at the University of Kansas, shown in Figure 2. It is equipped with a Ublox-Lea 6h global positioning system (GPS) receiver and a Paparazzi Tiny WithOut GPS (TWOG) autopilot, which can support both manual and autonomous flight. The detailed specifications are provided in Table 1. The KHawk UAS can carry one camera at a time for image acquisition. In this paper, two PeauPro82 modified GoPro Hero 4 Black cameras (PeauProductions 2017) are used for multispectral image acquisition with one providing imagery in the visible spectrum (red, green, blue (RGB)) and the other in the NIR spectrum. The spectral characteristics of the modified GoPro camera are shown in this section.

The KHawk 2 UAS was deployed for multispectral data acquisition over the study area on 7 June 2017. Two flights were conducted, one for NIR (09:49–10:19 A.M.) and the other for RGB (12:11–12:32 P.M.) video acquisition. It was flown autonomously at an altitude of 120 m above ground level with

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Takeoff weight (kg)</td>
<td>1.7</td>
</tr>
<tr>
<td>Wingspan (m)</td>
<td>1.2</td>
</tr>
<tr>
<td>Cruise speed (m/s)</td>
<td>16</td>
</tr>
<tr>
<td>Endurance (min)</td>
<td>30</td>
</tr>
<tr>
<td>Spatial resolution (at 120 m above the ground) (m)</td>
<td>0.1</td>
</tr>
</tbody>
</table>
a minimum horizontal and vertical overlapping percentage of 75% between flight lines to ensure accurate image orthorectification. The KHawk 2 flight locations (white circles) which are used for the orthorectification are shown in Figure 2. The acquired images were georeferenced and orthorectified using Agisoft Photoscan Professional software to produce orthomosaics in visible and NIR bands at a spatial resolution of 0.1 m. The spatial accuracy of the generated orthomosaics was calculated by performing a comprehensive comparison with ArcGIS World Image which has a spatial resolution of 0.5 m or better in the continental United States (ArcGIS 2021). A total of 10 high-quality control point pairs were randomly selected from the UAS and ArcGIS World images for comparison. The horizontal root-mean-square error (RMSE) for these 10 points is 4.72 m.

**LandSat 8 Satellite Data**

The L8 satellite is equipped with OLI and thermal infrared sensor instruments for multispectral image acquisition. The OLI measures light at nine spectral bands ranging from 430 to 1380 nm (NASA 2019). Operating at an altitude of 705 km above ground level, the OLI can provide calibrated reflectance images of the Earth every 16 days at spatial resolutions of 15 m (panchromatic band) and 30 m (all other bands). In this paper, OLI L2 SR images of the study area acquired on 7 June 2017 at 12:00 P.M. are used. The OLI L2 SR images are derived through the atmospheric correction of OLI Level-1 (L1) products using the LandSat Surface Reflectance Code algorithm (USGS 2020). The L8 OLI images demonstrate a spatial accuracy of 12 m or better (NASA 2021).

The KHawk GoPro and OLI sensor spectral characteristics (PeauProductions 2017) are compared and shown in Table 2. It can be observed that the NIR band of OLI is similar to the GoPro camera, while the red band of OLI is narrower than that of the GoPro camera.

Table 2. Spectral characteristics of KHawk 2 and Landsat 8 remote sensing systems.

<table>
<thead>
<tr>
<th>Sensor</th>
<th>Bands</th>
<th>FWHM (nm)</th>
<th>Peak (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLI</td>
<td>red</td>
<td>635.85–673.32</td>
<td>654.59</td>
</tr>
<tr>
<td></td>
<td>NIR</td>
<td>850.54–878.79</td>
<td>864.67</td>
</tr>
<tr>
<td>Modified PeauPro 82 GoPro (RGB)</td>
<td>red</td>
<td>583.9–710</td>
<td>627.6</td>
</tr>
<tr>
<td>Modified PeauPro 82 GoPro (NIR)</td>
<td>NIR</td>
<td>825.4–880</td>
<td>852.7</td>
</tr>
</tbody>
</table>

FWHM = full width at half maximum; OLI = Operational Land Imager; NIR = near-infrared; RGB = red, green, blue.

**Spectral Reflectance Estimation of UAS Imagery**

This section provides detailed descriptions of the proposed method for spectral reflectance estimation of multispectral UAS DN imagery. The objective of this method is to convert raw UAS images in DN at high spatial resolution to reflectance using available satellite reflectance data. The main advantage of this method is that spectral reflectance can be estimated from UAS DN images for free or at very low cost. In addition, this method will ensure similarity between the UAS and satellite imagery, which can later be used for multi-scale remote sensing applications of large fields such as crop monitoring and disaster assessment. The proposed method can be broken into the following steps for researchers who want to collect new UAS data.

1. **UAS campaign planning:** Plan the UAS data acquisition based on the availability of satellite data and favorable weather conditions. UAS campaign should follow the time schedule of satellite imagery collection.

2. **UAS data processing:** Collect UAS multispectral imagery of a given region of interest and perform georeferencing and orthorectification.

3. **Satellite data acquisition:** Satellite reflectance data of the region of interest is recommended to be acquired based on certain factors including acquisition time, cloud cover, and atmospheric correction (explained in further detail in the section “Satellite Data Acquisition”).

4. **Cross-calibration and conversion:** After the acquisition of UAS DN and satellite reflectance imagery, the SCC method can be implemented to identify the cross-calibration functions for each spectral band (explained in further detail in the section “Satellite-Based Cross-Calibration Method”). The identified cross-calibration functions can then be used to convert raw UAS DN imagery to reflectance values.

For researchers who have already collected their raw UAS data without ground calibration boards or ground spectrometer measurements, the following steps can be used instead.

1. **Find suitable reference of surface reflectance either from satellites (L8, Sentinel 2, Planet, etc.) or from airborne data (National Agriculture Imagery Program (NAIP), NEON, etc.).**

2. **UAS data processing:** (same as above).

3. **Cross-calibration and conversion:** (same as above).

**Satellite Data Acquisition**

Satellite reflectance data for a given region of interest (UAS survey area) can be downloaded for free or at low cost from designated websites such as the United States Geological Survey (USGS) Earth Explorer, Copernicus Open Access hub, Planet, etc. Reflectance data from any satellite (Landsat, Sentinel, Planet, etc.) can be used in this method and the following factors need to be considered during the selection.

1. **Acquisition time:** The satellite data acquisition time can be a critical factor in the effectiveness of the method. It is desired that the satellite data is acquired at the same time as that of the UAS image so that the reflectance is the same. In case of unavailability of such data, satellite data acquired close to the time of UAS campaign can be potentially used.

2. **Cloud cover:** Ideally, cloud-free satellite observations of the region of interest are desired. In cases with partial cloud obstruction, the unobstructed areas can be used to identify the cross-calibration functions which can then be used to calibrate the UAS data in cloud-obstructed areas. If there is a total cloud obstruction, satellite images acquired before or after the UAS survey can be used if available.

3. **Atmospheric correction:** Satellite images are required to be atmospherically-corrected to convert UAS DN to surface reflectance. Therefore, L2 data is recommended.

Satellite images acquired at the same time with UAS are preferred for this method. In cases when no suitable satellite images are available, aircraft imagery such as those from NAIP or NEON can also be used to serve as a reference reflectance image of the same area. The SCC method is described in the following subsection.

**Satellite-Based Cross-Calibration Method**

Given an orthorectified UAS image \( X \) in DN at high spatial resolution \((Km \times Kn)\) pixels and a satellite atmospherically-corrected reflectance image \( Y \) at medium spatial resolution \((M \times N)\) pixels, a cross-calibration function \( F(X) \) can be identified for each spectral band that can convert UAS images in DN at high spatial resolution to spectral reflectance. Here, \( k \) is the ratio between the spatial resolutions of satellite and UAS images which can be derived from the data set. For example, \( k = 30 \) if the spatial resolutions of satellite and UAS images are 30 m and 1 m, respectively. The main steps of this method (Figure 3) include:
Figure 1. UAS image resampling: Resample the high-resolution UAS image \((X')\) to a medium-resolution image \((X)\) to match the spatial resolution of the satellite image \((Y)\). Existing methods like nearest neighbor, bilinear, or bicubic methods can be used. Bicubic interpolation is used in this work.

2. Pixel selection: Select UAS and satellite pixel pairs at medium spatial resolution, \(X_i, Y_i\), which is a subset of the original UAS and satellite image pair, \((X, Y)\). Here, the objective is to exclude pixels that can potentially induce errors in the function identification and is explained in detail later in the section “Pixel Selection”.

3. Function identification: Use least-squares optimization methods to find the optimal cross-calibration function based on selected pixel pairs.

4. UAS reflectance estimation: Apply the identified function to the high-resolution UAS DN image \((X')\) and finally obtain UAS reflectance image \((Y')\).

**Pixel Selection**

In this work, the pixel pairs are selected based on two conditions, (a) low subpixel variability and (b) nonshadow pixels. The first condition is to exclude pixels with high subpixel variability within a medium-resolution pixel (e.g., a 30×30 m area for one OLI pixel). This is an important step as it can reduce the impact of mixed pixel effect of the lower resolution satellite image on the overall method. Essentially, this condition filters out the pixels that correspond to areas with varying reflectance values. This is determined based on the subpixel coefficient of variation (CV), calculated from high spatial resolution UAS image in DN. CV is the standard deviation divided by the mean value \((\sigma/\mu)\) within a pixel at medium-resolution. The concept is further shown in Figure 4. \(CV_g\) represents the CV of pixel \(B\) in \(X\), which is calculated using the mean \((\mu_{i})\) and standard deviation \((\sigma_{i})\) of all the pixels in a subgroup \(A(k \times k)\) in \(X\'). A threshold can be empirically defined to exclude all pixels in \(X\) with higher CV. The recommended threshold can be selected as the mean CV of all the pixels in the image, CV.

**Function Identification**

The objective of this step is to identify the cross-calibration functions between the selected sample, \((X_i, Y_i)\). Both linear and nonlinear functions have been used in the literature to convert DN to reflectance. In this paper, the exponential function is selected based on recent literature (Wang and Myint 2015; Deng et al. 2018) using similar cameras (commercial grade). The ordinary least squares (OLS) and weighted least squares (WLS) regression methods are used for the parameter identification of the optimal cross-calibration functions.

Ordinary least squares: The OLS method is used to estimate the unknown parameters in a linear regression model by minimizing the sum of squared errors. It is worth emphasizing that the OLS method is optimal under the assumption that the errors or the residuals are homoscedastic and serially uncorrelated across the measurement range (StatisticsSolutions 2020). For remote sensing images from different sources and at varying spatial resolutions, there is a high possibility of outliers and changing uncertainties across the reflectance range which may violate this condition.

Weighted least squares: When the OLS assumption of constant variance in the errors is violated (not homoscedastic), the WLS method can be used. The main difference is that in WLS, each data point in the given set of observations is weighed differently based on their resulting error variance. For example, data points with high error variances (e.g., outliers) are weighed very low compared to points with lower error variances. As a result, the WLS method is less sensitive to outliers as compared to the OLS method and has been used.
in remote sensing (Zhang et al. 2008; Shimabukuro and Smith 1991; Nencini et al. 2008). Another difference is that the WLS minimizes the sum of weighted squared error instead of sum of squared error as shown below (Mathworks 2020):

\[ S = \sum_{i=1}^{n} w_i (y_i - y_j)^2, \]

(1)

where \( w_i \) are the weights of each data point.

In this paper, a variant of the WLS, called the iteratively weighted least square regression (IWLSR), is used. This method has been successfully implemented for automatic relative radiometric normalization of satellite imagery (Zhang et al. 2008). The IWLSR method, initialized with the error residuals from the OLS method, iteratively determines the optimal weights for each data point in the given set of observations. The method is described below (Zhang et al. 2008).

Consider a set of observation data, \((x_j, y_j)\), such that \(j = 1, 2, 3,..., n\). The OLS is first used to estimate the slope and intercept, \(m\) and \(p\), respectively, such that:

\[ y_j = mx_j + p + \epsilon_j, \]

(2)

where \( \epsilon_j \) is the error residual for the \(j\)th data point. Next, a variable \( t \) is defined such that \( t = \epsilon_j/\sigma_\epsilon \), where \( \sigma_\epsilon \) is the standard deviation of the error residual vector. Here, \( t \) is an approximate one degree of freedom chi-square distribution \( (\chi^2(1)) \). This is used to calculate the weight vector, \( w \), for the next iteration.

\[ w = P(\chi^2(1) > t) = 1 - P(\chi^2(1) \leq t). \]

(3)

\( P \) in the above equation represents the chi-squared cumulative probability function. The above steps constitute one iteration. With the new weights, the updated slope and intercept are calculated, which starts the next iteration. The algorithm runs until the difference in the weights between two consecutive iterations falls below a specified value.

Finally, cross-calibration functions are identified for each spectral band that can be used to estimate spectral reflectance of UAS DN images. The SCC method is demonstrated in the next section.

Results

The results of the proposed SCC method are presented in this section using a multispectral data set, including KHawk UAS DN images and OLI SR data. The estimated KHawk high-resolution reflectance images are compared with high-resolution NIS SR images for validation.

Reflectance Estimation of KHawk UAS DN Images

The multispectral data set is shown in Figure 5, including orthorectified high-resolution (1 m) KHawk DN, resampled medium-resolution (30 m) KHawk DN, and medium-resolution (30 m) OLI SR images for the NIR band. The high-resolution and the resampled medium-resolution KHawk DN images share the same pixel range of 100–200 and the OLI SR image ranges from 0–0.6.

Pixel pairs are selected using the pixel selection method described in the section “Satellite Data Acquisition”. In this paper, the mean CV in the image \(X\), \( CV \) is used as the threshold to perform this selection. Figure 6 shows the selected pixels (circles) overlaid on the orthorectified and resampled UAS NIR DN image. As mentioned, the excluded pixels either represent high subpixel variability or shadows (example shown in Figure 6).

OLS and WLS regression methods are tested on the selected pixel pairs for function identification. Figure 7 shows the identified exponential cross-calibration functions using both methods for the NIR and red bands. Note that exponential
functions can be converted to linear functions by taking the natural log on both sides of the equation.

In order to identify the optimal exponential function from the aforementioned regression methods, error residuals and error variances are analyzed. Figure 8 shows the error residual plot for both methods, where the y-axis represents the error residual and x-axis \( y_{\text{hat}} \) represents the natural log of the estimated KHawk reflectance. It can be observed that the WLS method results in smaller and, more importantly, more consistent error variance as compared to OLS. The total error variance for NIR/red bands was 0.0061/0.0156 and 0.0011/0.0025 for the OLS and WLS methods, respectively. Therefore, the WLS method is selected for function identification. Table 3 shows the functions identified using the OLS and WLS methods. Here, \( x_i \) represents the KHawk DN values and \( y_i \) represents the OLI SR values with \( i = 1, 2, 3, ..., N \), where \( N \) is the total number of pixels in the selected pixel pairs. The identified functions are used to convert the high-resolution KHawk DN images to spectral reflectance for the NIR and red bands.

### Table 3. Identified functions using the OLS and WLS methods.

<table>
<thead>
<tr>
<th>Band</th>
<th>OLS</th>
<th>WLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIR</td>
<td>( y_i = 0.0483 e^{0.0135 x_i} )</td>
<td>( y_i = 0.0358 e^{0.0112 x_i} )</td>
</tr>
<tr>
<td>Red</td>
<td>( y_i = 0.0198 e^{0.00078 x_i} )</td>
<td>( y_i = 0.0160 e^{0.00095 x_i} )</td>
</tr>
</tbody>
</table>

OLS = ordinary least squares; WLS = weighted least squares; NIR = near-infrared.

Validation Using NIS Images

The estimated KHawk reflectance images at 1 m spatial resolution are compared to NIS SR images at the same resolution for validation. This subsection is broken down into: (1) NEON data description, (2) effects of spectral and spatial resolution on reflectance, and (3) NIS and KHawk reflectance comparison.

NEON Data Description

NEON is a continental-scale ecological observation facility project funded by NSF and operated by Battelle Memorial Institute (NEON 2019). NEON provides calibrated terrestrial, aquatic, atmospheric, and remote sensing data of 81 field sites across the United States to the scientific community, including the study area used in this paper. The NEON aircraft is installed with a pushbroom collection style NIS, which was designed and built by NASA’s Jet Propulsion Laboratory for hyperspectral remote sensing. The NIS measures light at 426 spectral bands (with 5 nm spectral resolution) ranging from 380 to 2500 nm and produces orthorectified reflectance mosaics at a spatial resolution of 1 m (NEON 2020). The aircraft is flown once each year over the study area at an altitude of 1000 m above ground level. The orthorectified images used in this paper correspond to a flyby on 9 June 2017 at 3:38 P.M. Note that these images are atmospherically corrected using the Atmospheric and Topographic Correction (ATCOR)-4 algorithm.

Effect of Spectral and Spatial Resolutions on Reflectance

Differences in spectral and spatial resolutions can have an effect on the changes in reflectance between images from different remote sensing platforms. In order to use high-resolution...
NIS images for validation, consistency in reflectance between NIS and OLI SR images needs to be established. NIS is a hyperspectral instrument and has a higher spectral resolution (narrower wavelength range) than the OLI. This can cause some inconsistencies between the reflectance images from both instruments. In order to minimize these differences, spectral convolution on the hyperspectral data can be performed to match the spectral response of the multispectral sensor (Zhao et al. 2010; Jarecke et al. 2001; Barry et al. 2002). In this paper, a weighted sum-based spectral convolution method is used, where the spectral resolution of the hyperspectral data is transformed to a lower value (Badawi et al. 2019; Meyer and Chander 2007). The weights are determined using the OLI prelaunch relative spectral response (RSR) for the NIR and red bands (NASA 2020). The equation used to perform the convolution is shown below:

$$\rho_i = \frac{\int_{\lambda_1}^{\lambda_2} \rho_i(\lambda) \cdot RSR_i(\lambda) \, d\lambda}{\int_{\lambda_1}^{\lambda_2} RSR_i(\lambda) \, d\lambda},$$

where $\rho_i$ is the spectrally convoluted NIS SR corresponding to band $i$, $\rho_i(\lambda)$ and $RSR_i(\lambda)$ are the NIS SR and the OLI spectral response of wavelength $\lambda$ in band $i$, respectively, and $\lambda_1$ and $\lambda_2$ are the lower and upper wavelengths of the OLI spectral range for band $i$.

In addition to spectral resolution differences, which are instrument-dependent factors, differences in spatial resolution (study area-dependent) can also cause reflectance inconsistencies between images from different platforms. Goetz (1997) observed that images of uniform grasslands show consistency across different spatial resolutions. Similar patterns have also been observed in open grasslands with varying grass species (Liu et al. 2017). Given these findings, the grass regions of the study area are expected to be consistent across different spatial resolutions. However, the tree regions may show differences due to factors such as canopy shadows and intertree spacing, which are generally observed in finer detail in high-resolution images.

The rest of this subsection is focused on analyzing the effect of spectral and spatial resolutions on reflectance for the study area, including (1) spectral resolution effect: comparison between resampled NIS and OLI images at 30 m resolution and (2) spatial resolution effect: comparison between high-resolution (1 m) and resampled medium-resolution NIS images (30 m) for the grass and tree regions. The high-resolution NIS and medium-resolution OLI NIR images are shown in Figure 9.

**Effect of Spectral Resolution:** Although the NIS images are spectrally convoluted using Equation 4, some differences are still expected. The NIS NIR and red SR images are resampled to 30 m resolution and compared to corresponding OLI images, shown in Figure 10. It can be observed that while the NIR band shows high similarities for most of the points, a constant bias is seen in the red band. The RMSE and mean percentage of the difference between images from the two sources are calculated for each band to be 0.0112 and −0.0043 for the NIR band and 0.0014 and −0.0274 for the red band.
Effect of Spatial Resolution: The differences between the high-resolution (original) and resampled NIS images are calculated for the grass and tree areas separately. One pixel in the resampled NIS image is compared to the mean of the corresponding 30×30 pixel grid in the high-resolution NIS image. RMSE and mean percentage error (MPE) are calculated for each band and shown in Table 4. Additionally, subpixel variability within each 30×30 pixel grid is also calculated using the CV. Mean CVs for grass and tree areas are also shown in Table 4.

Table 4. Effect of spatial resolution on reflectance in grass and tree areas.

<table>
<thead>
<tr>
<th>Region</th>
<th>Mean CV (%)</th>
<th>RMSE</th>
<th>MPE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>NIR</td>
<td>Red</td>
</tr>
<tr>
<td>Grass</td>
<td>11.01</td>
<td>17.17</td>
<td>0.028</td>
</tr>
<tr>
<td>Tree</td>
<td>30.85</td>
<td>44.97</td>
<td>0.040</td>
</tr>
</tbody>
</table>

CV = Coefficient of Variation; RMSE = root-mean-square error; MPE = mean percentage error; NIR = near-infrared.

It can be observed that the grass areas show higher consistencies and lower CV than the trees. The change of spatial resolution has a small impact on the reflectance in grass areas, which also provides a justification for our proposed cross-calibration method.

NIS and KHawk Reflectance Comparison

The estimated KHawk spectral reflectance images at 1 m spatial resolution are compared with NIS SR in detail in this section. First, the KHawk images are registered with the NIS images using a point mapping technique in MATLAB (MATLAB 2020). This step is necessary to reduce errors due to misalignment between the two images. A projective transformation is performed on the KHawk image using 12 selected control point pairs. The registered NIR and red reflectance images from both platforms are shown in Figure 11. Note that the black holes in the NIS images (close to the left boundary) and in the KHawk Red image (right boundary) are caused by orthorectification and are not used for comparison. Two analyses are performed for comparison (1) using a 32-subplot hayfield close to the northwest boundary and (2) using selected grass and tree regions.

Hayfield Comparison. The 32-subplot hayfield serves as a good reference for comparison because (1) the area is observed in images from both platforms and each subplot (grid size 10×10 m) is distinguishable, (2) the area is treated annually by KBS and the same treatment is given within each subplot, ensuring that the reflectance within each subplot is similar (Foster et al. 2010). Figure 12 shows the hayfield observed by the NIS (L) and KHawk (R) NIR images. The black box in each subplot represents a 6×6 m window. All pixels within the window are averaged to produce one value per subplot.
The averaged KHawk and NIS SR values for each subplot are compared in Figure 13. A good agreement can be observed between both images for the NIR and red bands with high correlations of 97% and 95.74% and low RMSE of 0.0239 and 0.0096.

**Selected Region Comparison.** In addition to the hayfield analysis, other regions in the area are also selected for NIS and KHawk comparison. It is worth mentioning that comparing all the pixels between the two images is difficult here due to pixel alignment and georeferencing uncertainties. In fact, KHawk orthorectified images are generated from many images and has a RMSE error of 4.72 m. Alternatively, six 3×3 m regions are manually selected for comparison, including three grass regions and three tree regions, shown in Figures 14 and 15. The region size (3×3 m) is selected based on the average tree canopy size observed in this data set. Note that shadows are excluded from the selected regions for a fair comparison. The six NIS and KHawk reflectance values and differences between them are shown in Figure 16 and Table 5, respectively. Mean absolute error and RMSE were found to be 0.0243 and 0.0306 for the NIR band, and 0.0178 and 0.0163 for the red band.

Two trends can be observed from Table 5: (1) tree regions exhibit higher differences in both NIR and red bands, and (2) red band shows slightly higher differences across most of the regions. These observations are similar to those found in the section “Effect of Spectral and Spatial Resolutions on Reflectance”.

In summary, the estimated KHawk reflectance images showed high similarities with the NIS images in grass regions, and slightly lower similarities in the tree regions. This shows the potential of the proposed SCC method for other similar fields (e.g., dominant grassland), given the availability of an atmospherically corrected satellite reflectance image.

**Discussions**

The SCC method presented in this work provides an effective and low-cost solution for the spectral reflectance estimation of UAS DN images, under the assumption that there exists satellite SR data of the same area around the same time. The main advantage of this method is that UAS images can be converted to reflectance images without using ground reflectance target boards and expensive spectroradiometers. Apart from its contributions to the UAS community, the SCC method can also greatly benefit the satellite remote sensing community by establishing cross-calibration functions between UAS and satellites, thus enabling future development of multi-source and multi-scale data fusion such as super-resolution for satellite imagery.

Since the proposed method uses images from different instruments, co-registration accuracy, differences in spectral resolution (Teillet et al. 1997), and conditions during data acquisition between UAS and satellite images can play a vital role in the accuracy of the function identification. Performing an accurate co-registration analysis between UAS and satellite...
images can be a challenging task considering the differences in their spatial resolutions (0.1 m and 30 m, respectively). However, since the spatial accuracy of the KHawk (4.72 m or better) and Landsat OLI (12 m or better) images are smaller than the pixel size of the Landsat OLI image, the impact of co-registration accuracy on the identified functions are not expected to be substantial. In this paper, images from three sensors are used, namely, KHawk 2 UAS GoPro, OLI, and NIS. Table 2 shows that the spectral characteristics between the GoPro and OLI sensors have small differences in peak wavelength (<27 nm) for both bands (Table 2). The NIS SR data is spectrally convoluted to match the spectral response of the OLI sensor using Equation 4. It can be observed from Figure 10 that although the NIR images from the NIS and OLI sensors show high similarities, the red images have some differences. Similar trends were also observed in Badawi et al. (2019), where the visible band images between NIS and OLI showed lower similarities. Additionally, the OLI and NIS SR data are subject to uncertainties of 5–10% respectively (Badawi et al. 2019), which can also cause differences between their respective SR images.

Conditions such as sun angle, clouds, and weather can also affect the accuracy of the proposed method. The ideal case is when the satellite and UAS images are acquired at the same time under similar weather conditions. However, due to practical issues, images acquired a few days apart are also acceptable, especially during the growing season (April–September), when the weather conditions are similar. The KHawk and OLI images used in this paper were acquired on the same day (7 June 2017) with the KHawk NIR and RGB images acquired during 9:49–10:19 A.M. and 12:11–12:32 P.M., respectively, and the OLI images at 12:00 P.M. The NIS image was acquired two days later, on 9 June 2017 at 3:38 P.M. It can be observed from Figure 14 that leaf canopy shadows are observed in opposite directions in NIS (towards east) and UAS (towards west) images. While the effect of shadows on the grass areas is not significant, more analysis is needed to understand why they affect the reflectance calculation for tree regions.

The proposed method is developed and tested on a forest/grassland area in Kansas, which is dominated by tall grass and is expected to show small variations in reflectance values with changes in spatial resolution. This is further validated by comparing high-resolution and resampled medium-resolution NIS SR images.

Figure 14. Selected tree regions in National Ecological Observatory Network Imaging Spectrometer (L) and KHawk (R) near-infrared reflectance images (3×3 m for each box).

Figure 15. Selected grass regions in National Ecological Observatory Network Imaging Spectrometer (L) and KHawk (R) near-infrared reflectance images (3×3 m for each box).

Table 5. Reflectance differences between NIS and KHawk.

<table>
<thead>
<tr>
<th>Region</th>
<th>NIS</th>
<th>KHawk</th>
<th>Difference</th>
<th>NIS</th>
<th>KHawk</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grass</td>
<td>0.2581</td>
<td>0.2595</td>
<td>0.0014</td>
<td>0.0577</td>
<td>0.0658</td>
<td>0.0082</td>
</tr>
<tr>
<td>Grass</td>
<td>0.2599</td>
<td>0.2736</td>
<td>0.0137</td>
<td>0.0487</td>
<td>0.0583</td>
<td>0.0096</td>
</tr>
<tr>
<td>Grass</td>
<td>0.2706</td>
<td>0.2787</td>
<td>0.0081</td>
<td>0.0501</td>
<td>0.0601</td>
<td>0.0100</td>
</tr>
<tr>
<td>Tree</td>
<td>0.4957</td>
<td>0.5454</td>
<td>0.0497</td>
<td>0.0343</td>
<td>0.0547</td>
<td>0.0204</td>
</tr>
<tr>
<td>Tree</td>
<td>0.5235</td>
<td>0.4762</td>
<td>-0.0473</td>
<td>0.0224</td>
<td>0.0476</td>
<td>0.0251</td>
</tr>
<tr>
<td>Tree</td>
<td>0.4860</td>
<td>0.5118</td>
<td>0.0258</td>
<td>0.0229</td>
<td>0.0473</td>
<td>0.0244</td>
</tr>
</tbody>
</table>

NIS = National Ecological Observatory Network Imaging Spectrometer; NIR = near-infrared.
other land cover types using ground spectroscopy measures the proposed method with more complex landscapes and for the tree regions.

0.0137 and 0.01 for the grass regions and 0.0497 and 0.0251 similarities are observed with reflectance differences less than by comparing with KHawk

2

grassland area, comprising of images from a low-cost KHawk posed method showed its effectiveness through a forest/
calibration targets or expensive spectroradiometers, which only utilizes publicly available data without using ground

In this paper, a low-cost and novel

Conclusion

Lastly, the proposed method has the following constraints based on the current study: (1) requirement of atmospherically-corrected satellite reflectance data which can be used as a reference for the UAS survey. Note that airborne reflectance data such as NAIP or NEON can also be used when satellite data is not available, (2) the accuracy of the reference data and their acquisition time will have an impact on the calculated UAS reflectance, and (3) the effectiveness of this method has been demonstrated in a grassland/forest field and further investigations are required to evaluate its performance in more complex landscapes and other land cover areas such as crop fields.

It is worth mentioning that the proposed method can be implemented using any UAS camera and satellite, including, Sentinel, Planet, etc. Also, the NIS images are used in this paper as a validation tool and is not required to implement the SCC method. In fact, this method showed its effectiveness in generating UAS reflectance image at 1 m, only because the available images for validation are at that resolution. It can potentially be applied to UAS images at higher resolutions (<0.1 m) also.

Future objectives include: (1) testing the effectiveness of the proposed method with more complex landscapes and other land cover types using ground spectroscopy measurements for validation; (2) development of machine learning-based algorithms for the mapping from UAS raw DN to reflectance and validation in different types of landscapes; (3) investigating the effect of BRDF correction on UAS reflectance estimation; and (4) comprehensive evaluation of this method on multiple data sets from different satellites such as Sentinel 2 and Planet.

Acknowledgments

The authors would like to thank Dr. Bryan Foster, Dr. Dean Kettle, Bruce Johanning, Sheena Parsons, Vaughn Salisbury, and Dr. Xueimin Tu from Kansas Biological Survey for their help with flight experiments and data collection. This work was partially supported by the NASA-KS-EPSRC grant NNX15AK36A R51438-2 and R51357-5, and Kansas Water Resources Institute USGS 104(b) Water Resources Research Grant 2018KS198B.

References


Early Classification Method for US Corn and Soybean by Incorporating MODIS-Estimated Phenological Data and Historical Classification Maps in Random-Forest Regression Algorithm

Toshihiro Sakamoto

Abstract

An early crop classification method is functionally required in a near-real-time crop-yield prediction system, especially for upland crops. This study proposes methods to estimate the mixed-pixel ratio of corn, soybean, and other classes within a low-resolution MODIS pixel by coupling MODIS-derived crop phenology information and the past Cropland Data Layer in a random-forest regression algorithm. Verification of the classification accuracy was conducted for the Midwestern United States. The following conclusions are drawn: The use of the random-forest algorithm is effective in estimating the mixed-pixel ratio, which leads to stable classification accuracy; the fusion of historical data and MODIS-derived crop phenology information provides much better crop classification accuracy than when these are used individually; and the input of a longer MODIS data period can improve classification accuracy, especially after day of year 279, because of improved estimation accuracy for the soybean emergence date.

Introduction

The population of the world is expected to increase to 9.7 billion by 2050 according to the United Nations (Department of Economic and Social Affairs 2019). Humans have to increase world food production by at least 25% in the next 30 y to meet the food demand of the additional 2 billion people. If the global balance of food supply and demand is destabilized because of global climate change in the future, developing countries that depend on food imports will become more vulnerable to volatility in international food prices, and food riots, such as those that occurred in Africa during the global food crisis of 2007–2008, can recur because of dramatic increases in food prices (Berazneva and Lee 2013). Against the backdrop of the situation surrounding international food supply and demand, a ministerial declaration of an Action Plan on Food Price Volatility and Agriculture was made at the meeting of G20 agriculture ministers in Paris in 2011. In the wake of this declaration, the Group on Earth Observations Global Agricultural Monitoring Initiative was launched to provide useful input through the use of modern tools, including remote sensing technology, to create an Agricultural Market Information System and internationally share accurate crop forecasts (G20 Agriculture Ministers 2011).

The United States is the world’s largest producer and exporter of agricultural products. Puma et al. (2015) have suggested that the global food system is vulnerable to systemic disruption in terms of connectivity and flows within the global food trade. Meanwhile, they point out that the US plays a large role in the global corn trade. For food-importing countries such as Japan, the importance of monitoring crop growth variation in the US using remote sensing technology is vital. For this reason, I devised a random-forest (RF) regression-based crop yield estimation method for US corn and soybean crops (Sakamoto 2020) by incorporating moderate-resolution imaging spectroradiometer (MODIS) time-series vegetation index data and meteorological and environmental data to reduce the prediction errors of a previous method (Sakamoto et al. 2014). The corn–soybean rotation is most common in major corn- and soybean-growing regions of the US. In regions where either crop cannot be grown, other upland crops—such as winter wheat, sorghum, beans, or alfalfa—are grown in combination with corn or soybeans. Because the same crop is not generally planted in the same field every year, a methodology for in-season crop classification is required to perform an essential function for the near-real-time crop-yield prediction system. Nevertheless, a classification algorithm for real-time crop-yield prediction for US corn and soybean crops at the MODIS pixel level has not been established yet. With this background, this study aimed to devise a simple and accurate classification method based on the RF regression algorithm.

Various approaches to crop classification have been proposed for crop monitoring at country and continental scales taking advantage of MODIS time-series data (Chang et al. 2007; Arvor et al. 2011; Conrad et al. 2011). Wardlow and Eggert (2008) applied a supervised decision-tree algorithm called See5 to the MODIS time-series 250-m normalized difference vegetation index (NDVI) data to produce three crop maps for Kansas: alfalfa, summer crops, and winter wheat. Shao et al. (2016) identified corn- and soybean-planted areas in the US from the multi-temporal NDVI by using a neural network classifier with the use of the cropland data layer (CDL) as training data in the Western corn belt. S. Zhang et al. (2019) developed a method combining observed crop phenology information and the MODIS time-series enhanced vegetation index to identify corn-planted areas in north and northeast China. Xun et al. (2018) applied sparse representation-based classification methods using 11 metrics derived from smoothed MODIS leaf area index data and meteorological and environmental data to produce three crops maps for China. However, the classification accuracy of these algorithms is not generally high enough for real-time crop classification.

The following conclusions are drawn: The use of the random-forest algorithm is effective in estimating the mixed-pixel ratio, which leads to stable classification accuracy; the fusion of historical data and MODIS-derived crop phenology information provides much better crop classification accuracy than when these are used individually; and the input of a longer MODIS data period can improve classification accuracy, especially after day of year 279, because of improved estimation accuracy for the soybean emergence date.
Materials and Methods

Study Area

The study area is focused on the five Midwestern US states of Illinois, Iowa, Kansas, Missouri, and Nebraska (Figure 1). In 2017, corn and soybean production in this area were 196 and 54.1 million tonnes, respectively, which accounted for 52.9% and 45.1% of the total national production (National Agricultural Statistics Service 2020). According to the Köppen–Geiger climate classification (Institute for Veterinary Public Health 2020), this area is located in regions of humid continental, humid subtropical, and semi-arid climate, covering diverse crop-growing environments. Corn is grown along the western border of Nebraska and Kansas, while soybeans are grown only in a few places west of approximately 100°W longitude. Annual precipitation decreases from the east to the west of the study area. While the growth limit for corn is 600 mm of annual rainfall in general, the boundary line of annual precipitation of 600 mm occurs between 98°W and 100°W (Figure 1). The irrigation farming area using groundwater from the Ogallala Aquifer is mainly spread west of this boundary line in the western region of Nebraska and Kansas (Pervez and Brown 2010), where winter wheat and sorghum are grown as rotation crops with corn.

MODIS Data

This study used the MODIS products of 8-d composite time-series data. The specifications of the MODIS products used in this study are summarized in Table 1. The data analyses for crop phenology detection and crop classification were performed at the original MODIS resolution of 231.7 m using sinusoidal projection. For the visualization of crop classification maps, the map projection was converted from the MODIS sinusoidal projection to equi-rectangular projection at a resolution of 250 m.

Table 1. Specifications of the data used.

<table>
<thead>
<tr>
<th>Data</th>
<th>Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>MODIS products</td>
<td>8-d composite surface reflectance (MOD09Q, MYD09Q, MOD09A, MYD09A)</td>
</tr>
<tr>
<td>Spatial resolution</td>
<td>231.7 m, 463.3 m</td>
</tr>
<tr>
<td>Spectral bands/layer</td>
<td>Band 1 (red) and band 2 (near-infrared) of MOD09Q and MYD09Q, band 3 (blue) and DOY layer of MOD09A and MYD09A</td>
</tr>
<tr>
<td>Map projection</td>
<td>Sinusoidal projection</td>
</tr>
<tr>
<td>Tile grids</td>
<td>h09v05, h10v04, h10v05, h11v04, h11v05</td>
</tr>
<tr>
<td>Period of data</td>
<td>From 2010 to 2019, DOY 01-365</td>
</tr>
<tr>
<td>Data distribution site</td>
<td>NASA Earthdata (<a href="https://earthdata.nasa.gov/">https://earthdata.nasa.gov/</a>)</td>
</tr>
</tbody>
</table>

Five data periods of different MODIS time-series data (Table 2) were set up to investigate the effect of the length of input data period on classification accuracy as a near-real-time simulation in the predictive experiment. Note that the meaning of
“day of year” (DOY) of the MODIS 8-d composite product name is the first date of the 8-d compositing period. For example, in the case of “MOD09A1.A2010201.h10v04.006.xxxx.hdf,” this composite product was made with observations from DOY 201 to DOY 208 of 2010. In this study, classification results are presented in terms of the date at which the predictive simulation results were available. In accordance with Sakamoto et al. (2014), the total time lag and waiting time were assumed to be 7 d from the last day of the MODIS 8-d composite data observation period through the product release on the NASA website to the end date of the classification analysis. When the classification analysis is performed using MODIS data, the input period is the shortest (DOY 65–201); the classification results will be available as of 3 August (DOY 215) during a non-leap year as the earliest predicted result under this assumption (Table 2).

Table 2. Input periods of the MODIS 8-d composite time-series data for crop classification methods.

<table>
<thead>
<tr>
<th>Notation in the Figures (Predictive Timing)</th>
<th>Input Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>DOY 215</td>
<td>DOY 65–201</td>
</tr>
<tr>
<td>DOY 247</td>
<td>DOY 65–233</td>
</tr>
<tr>
<td>DOY 279</td>
<td>DOY 65–265</td>
</tr>
<tr>
<td>DOY 311</td>
<td>DOY 65–297</td>
</tr>
<tr>
<td>DOY 343</td>
<td>DOY 65–329</td>
</tr>
</tbody>
</table>

**Cropland Data Layers**

This study used the CDL from 2008 to 2019. The latest data set is usually released in late January or early February on the National Agricultural Statistics Service website (https://www.nass.usda.gov/Research_and_Science/Cropland/SARS1a.php). The data sources of satellite images vary from year to year; the CDLs were made from optical sensor data with moderate spatial resolution including Landsat 5/TM, Landsat 8/OLI, ResourceSat-1/AWIFS, ResourceSat-2/LISS-3, Sentinel-2/MSI, and Disaster Monitoring Constellation Deimos-1 and UK2. The contiguous US was divided in the 2019 CDL into 115 classes of land use and land cover with 30-m resolution.

This study recategorized the CDL land cover/land use classes into three classes: corn, soybeans, and other. The mixed-pixel ratio of each class was calculated for each MODIS 231.7-m pixel on the condition that the total percentage of the three classes be 100%. Then the maximum value of the total mixed-pixel ratios of corn and soybean from 2008 to 2017 was calculated in each pixel to create a preliminary potential map of the region of interest for corn- or soybean-planted pixels, which was then used to limit the area of analysis and save calculation time (Figure 2). The areas in which the time-series maximum of the total mixed-pixel ratio of corn and soybean was at least 80% (Figure 2C) were defined as the target pixels to be analyzed, in accordance with criteria already defined (Sakamoto 2018). In preparing classification maps from the mixed-pixel ratios at 231.7-m resolution, the threshold used for classification of the target pixels as corn, soybeans, or other based on the estimated mixed-pixel ratio was 80%. If the mixed-pixel ratio of corn, soybeans, and other did not exceed the threshold, the pixel was classified into a fourth class named mixture.

**Refined-Shape Model-Fitting Method**

This study follows the concept of crop classification based on MODIS-derived crop emergence date (Sakamoto et al. 2014). The crop emergence date was estimated by the refined-shape model-fitting method (rSMF; Sakamoto 2018). The rSMF method explores the optimal values of three parameters that geometrically fit the preliminarily defined shape model, which was a smoothed profile of the wide dynamic range vegetation index (WDRVI; Gitelson 2004), on the MODIS WDRVI observations at the pixel level, on a near-real-time basis. Then the crop emergence date was estimated using the optimized values of two parameters with the preliminarily calibrated phenological parameters.
where \( X_n \) is the rSMF-estimated emergence date, \( x_{\text{scale}} \) and \( t_{\text{shift}} \) are the scaling parameters which geometrically fit the shape model on MODIS WDRVI time-series observations, and \( X_0 \) is the phenological parameter \( (X_0 = 147 \text{ in this study; for details, see Sakamoto et al. 2010; Sakamoto 2018}) \).

WDRVI has the advantage of higher sensitivity than NDVI to changes of leaf area index in moderate- to high-biomass conditions (Gitelson et al. 2007). It is calculated using red and near-infrared (NIR) reflectance, which is similar to NDVI except that it applies an additional weighting coefficient for the NIR reflectance:

\[
\text{WDRVI} = \frac{\rho_{\text{NIR}} - \rho_{\text{red}}}{\alpha \rho_{\text{NIR}} + \rho_{\text{red}}}
\]

where \( \rho_{\text{NIR}} \) and \( \rho_{\text{red}} \) are the MODIS 8-d composite surface reflectances in the NIR band (841–875 nm) and the red band (621–670 nm), respectively, and \( \alpha \) is a weighting coefficient (0.1 in accordance with Sakamoto 2018; this value was applied).

The maps of estimated crop emergence date as of DOY 215, 279, and 342 and the comparisons with the CDL are shown in Figures 3A2, 3B2 and 3E. In areas where corn–soybean rotation is commonly practiced, corn is generally planted earlier than soybeans, expecting to increase corn yield. Therefore, it can be assumed that the earlier emerging summer crop is corn and the latter is soybeans (Horvath et al. 1982), which agrees with observations of the National Agricultural Statistics Service of the United States Department of Agriculture (2020) indicating that the corn planting date is about 2 wk earlier than for soybeans in the homogeneous US corn belt. According to the crop progress records for the year 2017 provided by the United States Department of Agriculture Quick Stats Database (National Agricultural Statistics Service 2020), as of 21 May (DOY 141), 59% of Iowa’s cornfields were in or after the crop emergence period. Similarly, as of 4 June (DOY 155), 62% of Iowa’s soybean fields were in or after the crop emergence period.

**Procedures for Estimating the Mixed-Pixel Ratio for Crop Classification**

This study compares four classification methods. The process flowchart of the four methods using MODIS data and the CDL is shown in Figure 4. Table 3 shows a comparison of the explanatory variables and classification approaches of each method. The performances of the four methods were evaluated using leave-one-out cross validation (LOOCV) on the classification results from 2010 to 2017. In LOOCV, the data set was divided into 7 y of training data and 1 y of validation data. Then the division of this data set was repeated eight times to verify the performance of the proposed methods in 8 y. In addition, the classification accuracy of the proposed methods was evaluated assuming a near-real-time situation using data sets for 2018 and 2019, which were treated as completely independent test data from those of the LOOCV. Machine-learning models of the RF-based methods were separately created for each county.

**The Old Method Based on Emergence Date Without the RF Algorithm: Type 01 (Old Method)**

The first method, called Type 01 (Old method), is an improved version of the method of Sakamoto et al. (2014), which classified MODIS 231.7-m pixels into the three classes of corn, soybeans, or other (without including the mixture
class) by thresholding MODIS-estimated emergence date. The cropped area ratio between corn and soybeans within a county varied from region to region. Therefore, it was necessary to make a preliminary estimate of the county-level cropped area ratio between corn and soybeans using CDL data of the previous 2 y, to account for regional characteristics of the balance between corn and soybean acreage (Figure 4B). This ratio was used to define the threshold for splitting target pixels into corn- and soybean-planted pixels (Figure 5).

The details of the Type 01 (Old method) are as follows:

- Step 1: The potential area planted with corn or soybeans was identified by the historical maximum coverage area ratio of corn and soybeans (threshold: 80%; Figure 4A, Figure 2C).
- Step 2: The target pixels where the rSMF-estimated emergence dates were earlier than 1 March (DOY 61) or later than 1 September (DOY 244) were preliminarily classified as other. These thresholds (DOY 61 and 244) were empirically defined in terms of the improbable extreme emergence date of US corn or soybeans.
- Step 3: The target pixels excluding those classified as other were lined up in order of rSMF-estimated emergence date and then divided into two groups (earlier pixels as corn and later pixels as soybeans) by the threshold derived from the county-level area ratio between corn and soybeans (Figures 4B and 3).
- Step 4: Whereas the order of emergence dates was normalized from 0 to 100 by distinguishing between corn and soybeans (Figure 5), the normalized emergence date order of the other class was considered to be uniformly 100.

Then pseudo-composite color images were created by allocating these normalized emergence date orders of corn, soybeans, and other to RGB image layers (Figure 3C). The crop classification maps obtained are shown in Figure 3D.
The normalized emergence date order data for each class were used as explanatory variables for the Type 02 (EM) and Type 04 (Hybrid) classification methods, which were based on the RF algorithm.

The RF-Based Mixed-Pixel Decomposition: Type 02 (EM), Type 03 (CDL), and Type 04 (Hybrid)
As summarized in Table 3 and Figure 4, the RF-based methods (Type 02, 03, and 04) classified MODIS pixels into four classes (corn, soybean, other, and mixture) by thresholding the RF-estimated mixed-pixel ratio. The method called Type 02 (EM) also uses the normalized emergence date order as the explanatory variable, as in Type 01 (Old method); however, this method can estimate the mixed-pixel ratio of each class through the RF regression algorithm. The method called Type 03 (CDL) can estimate the mixed-pixel ratio with the RF algorithm but uses only the historical CDL data of the mixed-pixel ratio of each class for the previous 2 y. The method called Type 04 (Hybrid) can estimate the mixed-pixel ratio using both experimental variables used in Type 02 (EM) and Type 03 (CDL). This study investigated the effects of arbitrarily changing the thresholding value (which classified the target pixels) on classification accuracy and determined the optimum threshold to be used for the RF-based classification methods. Image processing of MODIS data, including map conversion, calculation of vegetation index, and the refined-shape model-fitting procedure, was done using ENVI version 5.6 and IDL version 8.8.0 (Harris Geospatial Solutions, Inc.). The program code for running the RF algorithm was written in Python (version 3.6.4, Anaconda custom 64-bit) installed on a Windows 10 PC. Version 0.19.1 of the machine-learning library scikit-learn was used for RF regression.

Results and Discussion

Time-Series Comparisons of Kappa Index for 2010–2017
The statistical values of the kappa index (median, maximum, and minimum) for the four methods for LOOCV from 2010 to 2017 are shown in Figure 6. Comparing between Type 01 (Old method) and Type 02 (EM), the kappa index commonly increased as the input MODIS data period became longer. It was greater than 0.25 after DOY 279 (Figure 6A and 6B). The degrees of classification agreement for Type 01 (Old method) and Type 02 (EM) were mid-fair (kappa index range: 0.21–0.40). Meanwhile, it can be seen that the yearly variation ranges of the kappa index for Type 02 (EM) were smaller than for Type 01 (Old method) even though both methods use the same explanatory variables of normalized emergence date order. The smaller range of kappa variation with Type 02 (EM) suggests that the addition of the RF regression algorithm had a positive effect on stabilizing classification accuracy.

The classification accuracy of Type 03 (CDL) was obviously better than that of Type 01 (Old method) and Type 02 (EM) even though it did not use any in-season satellite imagery. The kappa index of Type 03 (CDL) was 0.45 at a threshold of 65% (Figure 6C), which is in the first half of the moderate agreement level (kappa index range: 0.41–0.60). This implies that the machine-learning algorithm can predict the regional characteristics of the cropping pattern from historical data only with a certain degree of classification accuracy.

The Type 04 (Hybrid) method showed the best performance, with the highest kappa index at DOY 343 (median value: 0.57), which is 0.23 higher than Type 02 (EM) at a threshold of 60%, and 0.13 higher than Type 03 (CDL) at a threshold of 65%. The characteristics of the kappa index of Type 04 (Hybrid) reflected those of results from both Type 02 (EM) and Type 03 (CDL) (Figures 6B, 6C, and 6D). The kappa index of Type 04 (Hybrid) also gradually increased with the longer input period of MODIS data, as did the results of Type 02 (EM). Moreover, the kappa index of Type 04 (Hybrid) showed gentle arching shapes according to the threshold, which was close to the characteristic of Type 03 (CDL) rather than Type 02 (EM). These results imply that Type 04 (Hybrid) synergistically uses explanatory variables of the MODIS-derived phenological data and the historical CDL data. As a result, it had the highest classification accuracy (kappa: 0.51–0.58 at the threshold of 70%) among all the methods.

Estimation Accuracy of Mixed-Pixel Ratio
Figure 7 shows the 2D density plots of scatter distribution, which compares the mixed-pixel ratio of corn, soybeans, and other from the proposed classification methods and the ground-truth reference data (that is, CDL) using the single correlation coefficient r for the year 2017. The scatter distribution of the other class is concentrated especially near the bottom left corner of the density plots, where the mixed-pixel ratio ranges from zero to 25%, because most of the pixels in this class were preliminarily excluded in the process of making the region-of-interest maps of the target classes of corn and soybeans (Figure 4A). The numbers of pixels included in the three parts of the density plots (Figure 7D, parts 1–3) were compared to investigate the characteristics of each method in terms of mixed-pixel decomposition for corn and soybeans (Table 4). The three parts of focus were defined by the ranges of the mixed-pixel ratio of the density plots as follows:
- Part 1: Ground-truth reference data (CDL): 0%–25%; predicted values: 0%–25%.
Part 2: Ground-truth reference data (CDL): 75%–100%; predicted values: 75%–100%.

Part 3: Ground-truth reference data (CDL): 75%–100%; predicted values: 0%–75%.

It is noteworthy that the effects of the input MODIS data period remarkably increased the correlation coefficients of corn and soybeans from DOY 247 (corn: $r = 0.447$; soybeans: $r = 0.424$) to DOY 279 (corn: $r = 0.623$; soybeans: $r = 0.579$). This caused the kappa index of Type 02 (EM) to rapidly increase over the same period (Figure 6 B). It may be possible that Type 02 (EM) could accurately estimate subpixel coverages of corn and soybeans based only on the correlation coefficients after DOY 279 (corn: $r = 0.623–0.64$, soybeans: $r = 0.686–0.689$). Table 4 shows that for Type 02 (EM), from DOY 247 to DOY 279 for corn and soybeans, the number of part 2 pixels increased by 34.6% and 51.4% respectively and the number of part 3 pixels decreased by 24.4% and 22.6%. It was expected that the longer input MODIS data period would result in a better estimation of emergence date, especially for soybeans, which usually germinate after corn. The difference in the emergence date between corn and soybeans is more clearly visible on DOY 279 (Figure 3B2) than on DOY 215 (Figure 3B1).

Type 02 (EM) can roughly classify the target pixels into corn or soybeans; however, it cannot properly estimate the mixed-pixel ratio of each class. As can be seen from the density pattern of the scatter distribution (Figure 7A, Type 02), scatter plots are rarely distributed over the 1-by-1 straight line, especially from 25% to 75% (Figure 7D).

In contrast to Type 02 (EM), the scatter plots of Type 03 (CDL) are clearly distributed along the 1-by-1 straight line, unlike those with a higher correlation coefficient (corn: $r = 0.675$, soybeans: $r = 0.704$, other: $r = 0.651$; Figure 7, Type 03). This is probably because the RF model might evaluate whether each target pixel is located in the inside of the large field or at the boundaries of fields from the historical pattern of mixed-pixel ratio data for the previous 2 y, which consequently increases the estimation accuracy of the mixed-pixel ratios of each class in the next year. This shows that Type 03 (CDL) was superior to Type 02 (EM) only in terms of the correlation coefficient. According to Table 4, the number of

---

**Table 4. Number (thousands) of pixels in the three parts of density scatter plots comparing the mixed-pixel ratio derived by the prediction method and the reference crop classification map (CDL).**

<table>
<thead>
<tr>
<th>Crop</th>
<th>Type 02 (EM)</th>
<th>Type 03 (CDL)</th>
<th>Type 04 (Hybrid)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DOY 215</td>
<td>DOY 247</td>
<td>DOY 279</td>
</tr>
<tr>
<td>Corn</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Part 1*</td>
<td>1101</td>
<td>1162</td>
<td>1414</td>
</tr>
<tr>
<td>Part 2</td>
<td>801</td>
<td>840</td>
<td>1131</td>
</tr>
<tr>
<td>Part 3</td>
<td>1236</td>
<td>1198</td>
<td>906</td>
</tr>
<tr>
<td>Soybeans</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Part 1</td>
<td>1711</td>
<td>1724</td>
<td>2123</td>
</tr>
<tr>
<td>Part 2</td>
<td>466</td>
<td>475</td>
<td>719</td>
</tr>
<tr>
<td>Part 3</td>
<td>1092</td>
<td>1082</td>
<td>838</td>
</tr>
</tbody>
</table>

CDL = Cropland Data Layer.

*Each part of interest is shown in Plate 3D.

Mixture percentage range for each part of interest: Part 1—reference (CDL) = 0%–25%, prediction = 0%–25%; Part 2—reference (CDL) = 75%–100%, prediction = 75%–100%; Part 3—reference (CDL) = 75%–100%, prediction = 0%–75%.

---

- Part 2: Ground-truth reference data (CDL): 75%–100%; predicted values: 75%–100%.
- Part 3: Ground-truth reference data (CDL): 75%–100%; predicted values: 0%–75%.

It is noteworthy that the effects of the input MODIS data period remarkably increased the correlation coefficients of corn and soybeans from DOY 247 (corn: $r = 0.447$; soybeans: $r = 0.424$) to DOY 279 (corn: $r = 0.623$; soybeans: $r = 0.668$). This caused the kappa index of Type 02 (EM) to rapidly increase over the same period (Figure 6 B). It may be possible that Type 02 (EM) could accurately estimate subpixel coverages of corn and soybeans based only on the correlation coefficients after DOY 279 (corn: $r = 0.623–0.64$, soybeans: $r = 0.686–0.689$). Table 4 shows that for Type 02 (EM), from DOY 247 to DOY 279 for corn and soybeans, the number of part 2 pixels increased by 34.6% and 51.4% respectively and the number of part 3 pixels decreased by 24.4% and 22.6%. It was expected that the longer input MODIS data period would result in a better estimation of emergence date, especially for soybeans, which usually germinate after corn. The difference in the emergence date between corn and soybeans is more clearly visible on DOY 279 (Figure 3B2) than on DOY 215 (Figure 3B1).

Type 02 (EM) can roughly classify the target pixels into corn or soybeans; however, it cannot properly estimate the mixed-pixel ratio of each class. As can be seen from the density pattern of the scatter distribution (Figure 7A, Type 02), scatter plots are rarely distributed over the 1-by-1 straight line, especially from 25% to 75% (Figure 7D).

In contrast to Type 02 (EM), the scatter plots of Type 03 (CDL) are clearly distributed along the 1-by-1 straight line, unlike those with a higher correlation coefficient (corn: $r = 0.675$, soybeans: $r = 0.704$, other: $r = 0.651$; Figure 7, Type 03). This is probably because the RF model might evaluate whether each target pixel is located in the inside of the large field or at the boundaries of fields from the historical pattern of mixed-pixel ratio data for the previous 2 y, which consequently increases the estimation accuracy of the mixed-pixel ratios of each class in the next year. This shows that Type 03 (CDL) was superior to Type 02 (EM) only in terms of the correlation coefficient. According to Table 4, the number of
pixels for Type 03 (CDL) in the part 3 region of 2D scatter plots (915,000 for corn and 870,000 for soybeans) were not always lower than for Type 02 (EM) (906,000 for corn and 838,000 for soybeans as of DOY 279). This means that the producer accuracy of Type 03 (CDL) is slightly lower than for Type 02 (EM) on DOY 279. In other words, Type 03 (CDL) could not always more precisely classify corn- or soybean-planted pixels than Type 02 (EM), especially after DOY 279.

The density patterns for Type 04 (Hybrid) are extremely similar to those of Type 03 (CDL) in terms of accurate mixed-pixel decomposition. Furthermore, the correlation coefficient of Type 04 (CDL) is always higher than for Type 02 (EM) or Type 03 (CDL) after the beginning (Figure 7, Type 04). The number of pixels for Type 04 (Hybrid) in the part 3 region was lower than for Type 03 (CDL) on DOY 215 by 18.9% (corn) and 10.0% (soybean). Considering the difference in explanatory variables, it was expected that the effect of incorporating MODIS-derived phenological data into the RF algorithm would improve the estimation accuracy of Type 04 (Hybrid), especially for the part 3 region. Although Type 04 (Hybrid) showed the best performance, the predicted mixed-pixel ratio still tends to be too low in the part 3 region. This implies that the threshold for classifying target pixels into the dominant class or mixture class should be set to less than 80% to get higher classification accuracy. Considering the results shown in Figure 6, the threshold was fixed at 70% to create and validate classification maps for 2018 and 2019.

Classification Accuracy Assessment for 2018 and 2019

Figure 8 shows the crop classification maps for 2018 and 2019 as of DOY 343. Figure 9 shows the enlarged views of the crop classification maps as of DOY 215, 279, and 343 in the five distinctive environmental regions (areas 1–5 in Figure 8A1).

The target pixels are assigned two colors (red for corn or green for soybeans) even in regions where either corn or soybeans was grown along with other upland crops (Figure 8B). For example, in region 3 of Figure 8A2, Type 01 (Old method) could not properly classify the pixels, which actually were covered by winter wheat or sorghum, as other class (blue; Figures 8A1 and 8A2), misclassifying them instead as soybeans (green; Figures 8B1 and 8B2). Type 01 (Old method) suffers from the drawback of overestimating areas of corn- or soybean-cropped fields, especially in regions where
corn–soybean crop rotation is not common. The major improvements over Type 01 (Old method) in the other methods lead to proper detection of the distribution of pixels in the other class, especially in the central and western regions of Kansas and the western region of Nebraska (Figures 8C, 8D, and 8E). The classification maps derived from the other RF-based methods were quite similar in terms of overall spatial distribution patterns of each class. However, the differences in the characteristics of the results from each method are revealed by comparing the enlarged views (Figure 9).

In areas 1 and 2, where crop fields were cultivated with typical corn–soybean crop rotation, the results of Type 01 (Old method) and Type 02 (EM) do not show clear boundaries between corn and soybean fields on DOY 215 (Figure 9). However, the results are improved to be consistent with the true reference data (CDL) after DOY 279. Type 04 (Hybrid) shows more similarity in terms of spatial pattern to the true reference data (CDL) than Type 01 (Old method) or Type 02 (EM) from DOY 215 onward. As can be seen from the results for Type 03 (CDL), the rough pattern of the classification map can be predicted from the CDL data of the previous 2 y. The number of pixels in the mixture class for Type 04 (Hybrid) was lower than for Type 03 (CDL). This implies that Type 04 (Hybrid), with the additional use of normalized emergence date order, can more precisely improve the classification results to further categorize mixture pixels from Type 03 (CDL) into the corn or soybeans class. Both Type 01 (Old method) and Type 03 (CDL) were unable to classify the pixels in the other class (blue) properly. Type 02 (EM) could partially detect the pixels in the other class.

In area 3, where corn, soybeans, and other crops are grown in almost equal proportion (Figures 8 and 9), Type 01 (Old method) tended to misclassify pixels in the other class pixels as corn or soybeans, whereas Type 02 (EM) and Type 03 (CDL) tended to misclassify corn or soybean pixels as other or mixture. Although the classification results of Type 04 (Hybrid) included a certain number of mixture-class pixels that were not clearly determined as corn or soybeans, the overview of the spatial pattern was more consistent with the true reference data (CDL) than for other methods. The overall differences in performance observed in area 4 were similar to those for area 3.

In area 5, the results of Type 01 (Old method) were generally in good agreement with the true reference (CDL). However, the distribution area of the other class for Type 01 (Old method) tended to be smaller than the actual distribution. In terms of performance in detecting soybean-planted pixels, Type 04 (Hybrid) was superior to Type 02 (EM) and Type 03 (CDL), both of which could rarely detect them.

To summarize these visual assessments, the characteristics of each method are explained as follows. Type 01 (Old method) and Type 02 (EM) were unable to properly classify corn and soybeans by clearly distinguishing the boundary and shape of each field, especially in the earlier period of DOY 215. Type 03 (CDL) was good at roughly visualizing the shapes of crop fields and categorizing the pixels as either corn or soybeans, especially in typical corn–soybean rotation fields. However, it tended to categorize many target pixels or pixels of a regional minor class into the mixture class. Type 04 (Hybrid) showed the best performance by synergistically using the explanatory variables of normalized emergence date order and historical CDL data of the previous 2 y to visualize the shapes of crop fields and properly classify the target pixels from DOY 215 onward.

Assessment of Classification Accuracy in Terms of Geographic Characteristics

Figure 10 shows the accuracy verification results derived from Type 04 (Hybrid) for 2018 and 2019.

According to the state-level accuracy assessment, including kappa index, overall accuracy, and user and producer accuracies for corn and soybeans (Table 5), the classification accuracy was relatively higher in Iowa, Illinois, and Nebraska, especially after DOY 279. This is because the counties in which the kappa index was higher than 0.61 (moderate agreement)

![Figure 10](https://example.com/figure10.png)

Figure 10. County-level maps of kappa index and user and producer accuracies of the proposed Type 04 (Hybrid) method for 2018 and 2019.
spread out in a band from the eastern region of Nebraska to the central region of Illinois (Figure 10). This belt-like distribution area is consistent with the spatial distribution pattern of high-yielding counties (United States Department of Agriculture 2020), where both corn and soybeans are densely cropped (Figures 2A and 2B). The spatial variation of the user accuracy was less biased than that of the kappa index for both corn and soybeans in the study area (Figure 10). Many counties in Kansas, Missouri, and the western regions of Nebraska as well as Iowa showed a user accuracy > 70%, wherein the density of the corn- or soybean-planted area was higher (Figures 2A and 2B). Here, the user accuracy indicates how well the corn or soybean pixels identified by Type 04 (Hybrid) matched with the true reference data. If interpreted from the perspective of its use as part of a near-real-time crop yield-prediction system, the user accuracy can be considered the screening accuracy of the ability of Type 04 (Hybrid) to precisely extract target crop pixels. Counties that had a higher user accuracy showed less misclassification, which had a good effect on yield prediction results. The producer accuracy quantifies the corn or soybean pixels in the true reference data that were identified by Type 04 (Hybrid). In this case, the producer accuracy can be considered as an indicator of how many of the total true target pixels were identified by Type 04 (Hybrid). A higher producer accuracy indicates higher representativeness of pixels identified by Type 04 (Hybrid) to the whole set of true pixels within a county or state. The spatial patterns of the producer accuracy indicate a similar spatial bias to that of the kappa index, which would be related to the cultivation density per unit area.

**Discussion on Using Low-Resolution Satellite Data for Crop Classification**

Considering the differences in the spatial variations between the user and producer accuracies, it is evident that the proposed Type 04 (Hybrid) method can precisely detect corn- or soybean-planted pixels with less mixed-pixel effect in major cropping regions. This study reveals that the regional characteristic of classification accuracy depends on crop cultivation density. In that sense, the study can be expected to improve the accuracy of a near-real-time crop yield prediction model at the state level, if the results of crop yield predictions are scaled up from the county level to the state level, through a reliability-weighting technique based on the accuracy of the county-level classification. Conversely, it cannot detect pixels with a similar accuracy in regions with low crop density, because of the use of low-resolution satellite data. There is also the possibility of misclassifying crops with similar growing periods. Therefore, a limitation of the proposed method is that it cannot avoid errors in yield prediction caused by the mixed-pixel effect. In addition, it is important to note that extreme weather caused flooding especially along the Mississippi and Missouri Rivers in the Midwestern US due to heavy rains during late May and early June 2019 (National Weather Service 2019). The impact of the 2019 flooding on crop fields was confirmed in regions 1 and 2, where the croplands were classified as other (blue), whereas they were classified as “Fallow/Idle cropland” in the CDL-derived true reference map (Figure 8A2). This suggests that the remote sensing-based method could not properly classify crop fields that were planted with corn or soybeans in the spring and later abandoned by farmers for some reason by using any of the classification methods without collecting detailed ground-truth data.

In the past, the use of higher-resolution satellite images for continental-scale land use classification was economically unreasonable in terms of analysis using a personal desktop computer. The present study avoided increasing the computation time by using high-resolution satellite data to prioritize obtaining a wide-area crop classification map for quick reporting through a remote sensing-based early crop-yield prediction system. However, with the rapid expansion in cloud computing services such as Google Earth Engine and Amazon Web Services, the use of higher-resolution satellite images is becoming economically possible. Many studies have adopted a data-fusion approach that combines MODIS time-series data with higher-spatial-resolution data such as Sentinel-2/MSI and Landsat series/OLI, MSI, and MSI (Hilker et al. 2015; Jia et al. 2014; Li et al. 2015; Xiong et al. 2017; Onojeghuo et al. 2018; Dao et al. 2019; Qi and Wang 2019; Wang et al. 2020; Zhou et al. 2020), which is the easiest way to minimize the effects of mixed pixels on crop classification accuracy. In the near future, research applying the data-fusion approach on a cloud computing platform will become the standard for providing continent-scale detailed data on crop classification and phenology.

**Conclusions**

This study proposed crop classification methods to categorize US agricultural fields into four classes (corn, soybeans, other, and mixture) in near real time at 250-m resolution. The four
methods were compared to investigate the effects on classification accuracy of two types of explanatory variables and the machine-learning algorithm introduced. The previous method classifies pixels by thresholding the MODIS-estimated crop emergence date. Improvements to this approach were made by incorporating the RF regression algorithm to estimate the mixed-pixel ratio of each class using CDL data of the past 2 y as an explanatory variable in addition to the MODIS-derived phenological data. A comparison of the results of mixed-pixel ratios from the RF-derived estimation and the true reference data (CDL) shows that the RF-based methods tended to underestimate the mixed-pixel ratio of corn and soybeans within a MODIS pixel. This study applied a threshold of 70%, which showed the best classification accuracy on DOY 215, to investigate the functional characteristics of the proposed methods in terms of crop classification accuracy.

Although the Type 02 (Em) method could estimate the mixed-pixel ratio using the RF regression algorithm along with the normalized crop emergence date order, its classification accuracy—although more stable than that of the conventional Type 01 (Old method) approach—was not obviously improved. The Type 03 (CDL) method, incorporating only the CDL data of the previous 2 y as an explanatory variable in the RF regression algorithm, precisely estimated the mixed-pixel ratio of each class and performed better than Type 02 (Em) in terms of creating the crop classification maps with clear boundaries of crop fields. However, Type 03 (CDL) still tended to categorize many objective pixels as mixture, especially in the regions outside the main corn and soybean cropping area, where the three target classes (corn, soybeans, and other) were sparsely distributed together. The Type 04 (Hybrid) method, which uses both explanatory variables of the CDL data of the previous 2 y and the normalized emergence date order, showed the best performance in terms of estimating the mixed-pixel ratio and precisely classifying crop cover at the earliest timing of DOY 215. The accuracy assessment of Type 04 (Hybrid) for 2018 and 2019 was as follows: The kappa index ranged from 0.50 to 0.58. The overall accuracy ranged from 65.3% to 72.3%. The user accuracies for corn and soybean ranged from 72.2% to 77.3% and from 69.2% to 79.5%, respectively. The producer accuracies ranged from 70.0% to 83.2% and from 60.5% to 79.3%. The classification accuracy varied from region to region, and tended to be better in the major cropping regions of Iowa, Illinois, and Nebraska, which is related to the cultivation density of corn or soybeans per unit area. The Type 04 (Hybrid) method displayed advantageous characteristics of both Type 02 (Em) and Type 03 (CDL). The results with Type 04 (Hybrid) clearly revealed the field contour shape from DOY 215 onward because of better estimation accuracy of the mixed-pixel ratio than Type 02 (Em), which considered the effect of CDL data of the previous 2 y. Furthermore, the classification accuracy of Type 04 (Hybrid) increased with the longer input data period of MODIS time-series data from DOY 215 to DOY 343 because of the improvement in accuracy estimation of mixed-pixel ratios that was afforded by using the crop emergence date as an explanatory variable in the RF regression. Therefore, Type 04 (Hybrid) can create the most stable and precise classification results through the synergetic effect of crop discrimination ability based on RSIMF-estimated crop emergence date and the ability to predict future crop cover based on the historical CDL data. The salient conclusions of this study are as follows:

- The use of the RF regression algorithm is effective and convenient for estimating the mixed-pixel ratio using a low-resolution satellite image by incorporating different types of explanatory variables.
- The fusion of historical CDL data and MODIS-derived crop phenology information can improve crop classification accuracy compared to the use of each individually.
- The input of a longer MODIS data period improves crop classification accuracy, especially after DOY 279, because of better estimation of the emergence date, particularly for soybeans.

In a future study, I would like to integrate the proposed classification method with the improved crop-yield estimation model I devised previously (Sakamoto 2020) to realize a near-real-time crop monitoring system, which would operationally predict yields of corn and soybeans covering the entire United States of America.

Acknowledgments

I appreciate the valuable comments and suggestions of the anonymous reviewers, which gave me an opportunity to further improve the manuscript. This study was supported by JSFS KAKENHI (grantJP17K08037).

References


A Deep Multi-Modal Learning Method and a New RGB-Depth Data Set for Building Roof Extraction

Mehdi Khoshboresh-Masouleh and Reza Shah-Hosseini

Abstract
This study focuses on tackling the challenge of building mapping in multi-modal remote sensing data by proposing a novel, deep superpixel-wise convolutional neural network called DeepQuantized-Net, plus a new red, green, blue (RGB)-depth data set named IND. DeepQuantized-Net incorporated two practical ideas in segmentation: first, improving the object pattern with the exploitation of superpixels instead of pixels, as the imaging unit in DeepQuantized-Net. Second, the reduction of computational cost. The generated data set includes 294 RGB-depth images (256 training images and 38 test images) from different locations in the state of Indiana in the U.S., with 1024 × 1024 pixels and a spatial resolution of 0.5 ft that covers different cities. The experimental results using the IND data set demonstrates the mean F1 scores and the average Intersection over Union scores could increase by approximately 7.0% and 7.2% compared to other methods, respectively.

Introduction
Building monitoring can play a crucial role in Earth observation and environmental applications, such as urban planning and management, change detection, three-dimensional (3D) building reconstruction, infrastructure development, and many urban development and reconstruction projects (Akbulut et al. 2018; Mousa et al. 2019; Singh et al. 2015; Wang et al. 2016). Automatic building segmentation is one of the most critical and challenging issues in building monitoring (Shi et al. 2018; Wu et al. 2019). Building segmentation means that buildings are found, based on pixel-wise labeling, so each pixel is labeled with given concepts (Masouleh and Sadeghian 2019). Over the past few decades, several methods have been proposed for automatic building segmentation using remote sensing red, green, blue (RGB) images (Bayanlou and Khoshoresh-Masouleh 2020; Gu et al. 2018; Khoshoresh-Masouleh et al. 2020; WuDunn et al. 2020; Zhang et al. 2018). One of the simplest ways to automate building segmentation is the use of texture features extracted from grey level co-occurrence matrices to improve classification accuracy (Odoo 1992; Salhi et al. 2019). Moreover, machine learning-based methods, such as the support vector machine and the random forest, have greatly improved building segmentation (Gavankar and Ghosh 2018; Wicht and Kuffer 2019; Zheng and et al. 2020). Recently, deep learning (DL)-based algorithms in digital image processing have been vastly developed (Bi et al. 2020a, 2020b, 2020c; Jin et al. 2020; Khoshoresh-Masouleh and Shah-Hosseini 2020; Maier et al. 2019; Mohammedaslan and Uğuz 2020; Zhang and Chen 2020). The convolutional neural network is an essential model in deep learning for semantic segmentation. Two-dimensional (2D) convolutional neural networks use 2D convolutional kernels to predict a single channel image (cf. Figure 1). Because 2D convolutional neural networks take a single channel image as input data, they inherently fail to leverage the context of target-related information for segmentation from adjacent channels. 2D convolutional layers in the neural network can leverage the context across the surface area (height and width) of a single channel image to make predictions (Hu et al. 2019).

Figure 1. Two-dimensional convolution operation. The values of the input channels are multiplied with the convolution operator and summed together to form the output value (two-dimensional feature). For interpreting the references to color in this figure legend, the reader is referred to the section “Data Availability.”

DL models, by extracting the useful features of the image, can significantly improve building segmentation accuracy from multi-modal remote sensing data (Tripodi et al. 2019). In this regard, the fusion of RGB and depth data in DL models is an essential problem for multi-modal semantic segmentation (Koppanyi et al. 2019; Sefercik et al. 2014; Uzar 2014). Although the related DL algorithms are powerful, there is still not robust performance for building segmentation from multi-modal data acquired from different sensors, such as RGB and depth data, particularly in automatic building segmentation without any further postprocessing and handcrafted knowledge (Sun et al. 2019). Additionally, most of these algorithms are based on the basic model and just use pixels for processing (Khoshoresh-Masouleh and Hasanlou 2020; Xu et al. 2018). To achieve robust performance, the DL algorithm should give special attention to the processing unit of the input.

The purpose of this study is to develop and evaluate a new DL model aimed at improving building segmentation from multi-modal remote sensing data with a focus on tackling the problems of previous research. The proposed model is developed using vector quantization, superpixel, and atrous convolutional layers. Vector quantization is a data coding method.
by forming vectors such that the source output (e.g., image) will be grouped into vectors. The most important features of the vector quantization method are memory usage reduction while maintaining all the necessary data. Popular vector quantization algorithms include the tree-structured model (Gray et al. 1992), classified model (Chen et al. 2014; Wang and Yang 2020), and adaptive model (Gersho and Gray 1992). Here, we use a classified structure which is especially useful in image data processing for the deep competition model approach.

The proposed architecture combines different size atrous convolutional subblocks as the encoder stage and six transposed convolutional subblocks as the decoder stage. The design logic of this model is based on the optimization of extraction accuracy and processing time by combining vector quantization and atrous convolutional layers in the form of a convolutional neural network with an encoder-decoder format. In a convolutional neural network, the encoder-decoder structure is used to extract high-level features from image data. The encoding stage with atrous convolutions generates high-level features with lower spatial dimensions rather than the input dimension. The decoding stage with transposed convolutions is the reverse of the encoding stage to produce the final map. More precisely, the unique contributions of this paper are:

1. A fast and robust dilated dilated model was developed to estimate building footprints based on learning vector quantization theory and superpixels using multi-modal remote sensing data. This paper proposes a novel regularization strategy based on quantization error for the training of the proposed dilated model.
2. A new open-access multi-modal data set, called IND, for building extraction is proposed, which brings main challenges, including shadows and occulted areas, vegetation covers, complex roofs, and dense building areas.
3. A comprehensive evaluation based on the most critical challenges for building extraction from multi-modal data is presented.

The rest of this paper is organized as follows. In the section “Materials and Methods,” materials and methods present the details of the proposed DeepQuantized-Net and data sets. Next, in the section “Results and Comparisons,” we evaluate the performance of DeepQuantized-Net on a challenging building data set. The section “Ablation Experiments” is the description of the ablation experiments, and the “Discussion” section discusses the experimental results. Finally, the “Conclusions” section concludes this paper.

Materials and Methods

DeepQuantized-Net Architecture
In this study, an end-to-end trainable quantized deep superpixel-wise convolutional architecture (hereafter called DeepQuantized-Net for short) for building segmentation from multi-modal data is proposed. Additionally, three other state-of-the-art dilated image segmentation architectures, including residual network (Res-U-Net) (Xu et al. 2018), adaptive bilateral filter + segment-based neural network (ABF+SegNet) (Masouleh and Shah-Hosseini 2018), and deformable convolutional networks (DCN) (Khoshboores Masouleh and Saradjian 2019), were used for building segmentation results comparisons. All models were selected because they have effective performance in building segmentation from multi-modal remote sensing data.

DeepQuantized-Net incorporates two influential concepts in semantic binary segmentation. They are building segmentation and optimizing the weights of a deep encoder-decoder model with modeling competition between input and output vectors in the proposed network. In this competition, less expensive computation and higher accuracy (compared to the other models) are used. In the competition model, instead of pixels, a superpixel is used as the primary data unit to improve the object pattern vector. Figure 2 shows the structure of the proposed model for building segmentation based on the concepts described above. DeepQuantized-Net model is an encoder-decoder architecture built up of six encoder-decoder subblocks based on deep competitive learning. Deep competitive learning is the novel concept for convolutional encoder-decoder architecture.

Deep competitive learning improves the representation or feature learning by limiting the infinite input space using the integration vector quantization (Kohonen 1995) and convolutional layers. Since this building segmentation process is binary, joint vector quantization, and convolutional layers improve detection and segmentation procedures significantly. The quality of results of the DeepQuantized-Net model depends on modeling competition between input (x) and output (y) vectors. The competition model is defined as follows:

\[
\hat{k} = \arg\max_k \left\| \sum_{i=1}^{k} \frac{1}{1 + e^{x_i-y_k}} \right\|_{k=0,1} \tag{1}
\]

where \(x\) is the input superpixel; \(y\) is the output map; \(e\) is the Napier’s constant, and \(\hat{k}\) is the binary value (0 = nonbuilding, 1 = building).

The integral equation for the proposed competition model based on input superpixels probability distributions (P) can be defined as follows:

\[
E = \int \frac{1}{1 + e^{x-y}} \times P(x) \, dx \tag{2}
\]

where \(E\) is the quantization error in the competition model.

The overall scheme of the learning vector quantization theory based on superpixels in DeepQuantized-Net consists of the following:

1. Divide the multi-modal data segmented into homogeneous superpixels and feature extraction using convolutional filters with different dimensions.
2. Reconstruct the high-level patterns based on the decoding stage for classifying the feature vectors into the building and nonbuilding patterns.
3. Cluster the patterns of building and nonbuilding into binary classes by using the competition model (Equation 1). DeepQuantized-Net is an effective learner for spectral-spatial relationships of the multi-modal data, with a high level of abstraction using convolutional filters with different dimensions (e.g., \(3 \times 3, 5 \times 5, \) and \(7 \times 7\)), in the encoding stage for better separation of the marginal details of building footprints from other terrestrial objects. In the DeepQuantized-Net, the encoding block takes an input consisting of superpixels in RGB-depth and generates a high-level representation vector with aggregate characteristics.

In each subblock, a batch normalization (BN) function (Laurent et al. 2018) and quantization error-based regularization technique have been used to improve performance in the training stage with a focus on reducing overfitting. The activation function of this model is the Leaky Rectified Linear Unit (LReLU) function (Liu et al. 2020). In Xu et al. (2015), it is pointed out that the LReLU consistently outperforms the original Rectified Linear Unit. BN and LReLU functions are computed as follows, respectively:

\[
I_{BN} = \frac{I - B_n}{B_v} \tag{3}
\]

\[
\text{LReLU}(I) = \max(a, I) \tag{4}
\]

where \(I\) is the input layer, \(B_n\) is the batch mean, \(B_v\) is the batch variance, and \(a\) is a small constant.
Implementation

The building segmentation steps of this method are as follows: first, a superpixel generator approach called Simple Linear Iterative Clustering (SLIC) (Achanta et al. 2012), which is the robust superpixel segmentation method, is used to generate the main image unit in the proposed method.

SLIC is a local version of k-means clustering in visually homogeneous regions to decompose an image. Figure 3 shows the overview of RGB-D superpixel generation based on the SLIC method. In this study, RGB and depth information are combined to establish a multi-modal superpixel for building extraction. In general, the implementation levels of this method are fivefold:

1. Convert the RGB image (or three depth channels via copy) to CIELAB color space.
2. Initialize cluster centers by sampling pixels at regular grid steps.
3. Move cluster centers to the lowest gradient position in a 3×3 neighborhood.
4. Set label for each pixel.
5. Set distance for each pixel.

Second, the proposed model called DeepQuantized-Net is designed and implemented in Google Colab and Google Drive cloud computing environments. In the design of DeepQuantized-Net, a rectified adaptive moment estimation (RADAM) optimizer is used to optimize the parameters. As Xiang et al. (2020) pointed out, the RADAM consistently outperforms the original adaptive moment estimation. Third, the efficiency of the proposed method is evaluated compared to similar methods using ground truth maps and accuracy analysis measures.

(a) Superpixels of RGB Image

(b) Superpixels of Depth Image

Figure 2. DeepQuantized-Net architecture. For interpreting the references to color in this figure legend, the reader is referred to the section “Data Availability.”

Figure 3. An overview of red, green, blue-depth superpixels. For interpreting the references to color in this figure legend, the reader is referred to the section “Data Availability.”
Data Set
In this study, we use the IND building scenes data set to benchmark the performance of the proposed model. The IND data set contains 256 training and 38 test remote sensing RGB-D and pixel-level labeled images at 1024 × 1024 resolution, which brings new challenges, including shadows and occluded areas, vegetation covers, complex roofs, dense building areas, and large-scale variation. Additional information about data acquisition and data annotation are in the section “Supplementary Information” (cf. Figures S1 and S2). DeepQuantized-Net training has a learning rate of 0.01, a momentum factor of 0.9, and a batch size of 20.

Results and Comparisons
Experimental Results
To evaluate the DeepQuantized-Net model, eight test images were selected from study areas. The main features of these test images include: (1) high density of buildings and (2) different building sizes and textures. The best-trained model can be extracted with minimum validation loss (Khoshboresh Masouleh and Shah-Hosseini 2020, 2019). The best model for the IND data set is obtained at iteration 150. Results of the proposed model were compared with Res-U-Net, ABF+SegNet, and DCN methods, which are among the most essential DL methods.

Table 1 shows the visualized results for the DeepQuantized-Net model used in the eight test images from the IND data set. The results of the DeepQuantized-Net model indicate improvement in building segmentation. Based on the results of this study, using the filters with different dimensions in the proposed model has the following advantages: First, the proposed model generates new, high-level features and integrates them along with the architecture which is an appropriate approach for reinforcing the data. Second, access to global-local features is provided simultaneously. Additional qualitative results are in the section “Supplementary Information” (cf. Figure S3).

The average F1 score and Intersection over Union (IoU) score for complex scenes (e.g., samples 2, 4, and 5) are about 94.9% and 90.6%, respectively, while these values for different color roofs and directions scenes (e.g., samples 1 and 6) are about 98.3% and 96.6%, respectively. Based on Table 1,

Table 1. Comparison of performance of residual network (Res-U-Net) (RU), adaptive bilateral filter + segment-based neural network (ABF+SegNet) (AS), deformable convolutional networks (DCN), and DeepQuantized-Net (DQ) in the IND data set. The italic values denote the best result achieved by methods.

<table>
<thead>
<tr>
<th>Test</th>
<th>Metric</th>
<th>RU (%)</th>
<th>AS (%)</th>
<th>DCN (%)</th>
<th>DQ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F1</td>
<td>92.3</td>
<td>90.7</td>
<td>93.1</td>
<td>97.7</td>
</tr>
<tr>
<td></td>
<td>IoU</td>
<td>84.8</td>
<td>83.6</td>
<td>89.4</td>
<td>95.6</td>
</tr>
<tr>
<td>2</td>
<td>F1</td>
<td>81.5</td>
<td>81.8</td>
<td>83.2</td>
<td>89.1</td>
</tr>
<tr>
<td></td>
<td>IoU</td>
<td>69.6</td>
<td>69.5</td>
<td>71.4</td>
<td>80.3</td>
</tr>
<tr>
<td>3</td>
<td>F1</td>
<td>84.7</td>
<td>86.2</td>
<td>90.2</td>
<td>96.3</td>
</tr>
<tr>
<td></td>
<td>IoU</td>
<td>79.2</td>
<td>80.1</td>
<td>89.3</td>
<td>92.8</td>
</tr>
<tr>
<td>4</td>
<td>F1</td>
<td>90.4</td>
<td>91.7</td>
<td>92.4</td>
<td>97.4</td>
</tr>
<tr>
<td></td>
<td>IoU</td>
<td>85.8</td>
<td>88.2</td>
<td>89.8</td>
<td>94.9</td>
</tr>
<tr>
<td>5</td>
<td>F1</td>
<td>81.5</td>
<td>84.2</td>
<td>89.9</td>
<td>98.4</td>
</tr>
<tr>
<td></td>
<td>IoU</td>
<td>78.2</td>
<td>83.1</td>
<td>88.6</td>
<td>96.5</td>
</tr>
<tr>
<td>6</td>
<td>F1</td>
<td>85.1</td>
<td>85.5</td>
<td>90.3</td>
<td>98.9</td>
</tr>
<tr>
<td></td>
<td>IoU</td>
<td>83.8</td>
<td>84.2</td>
<td>89.1</td>
<td>97.6</td>
</tr>
<tr>
<td>7</td>
<td>F1</td>
<td>79.8</td>
<td>85.4</td>
<td>87.0</td>
<td>96.3</td>
</tr>
<tr>
<td></td>
<td>IoU</td>
<td>77.9</td>
<td>81.6</td>
<td>83.4</td>
<td>94.2</td>
</tr>
<tr>
<td>8</td>
<td>F1</td>
<td>86.1</td>
<td>88.4</td>
<td>90.2</td>
<td>97.7</td>
</tr>
<tr>
<td></td>
<td>IoU</td>
<td>83.6</td>
<td>85.6</td>
<td>89.4</td>
<td>96.6</td>
</tr>
<tr>
<td>T(min)</td>
<td>–</td>
<td>522</td>
<td>580</td>
<td>549</td>
<td>311</td>
</tr>
</tbody>
</table>

F1 = average score; IoU = Intersection over Union.
the average F1 score and IoU score for samples 3, 4, and 7 are 96.7% and 93.9%, respectively.

The experimental results can highlight the abilities of the proposed method to extract buildings from RGB-D data with the average F1 scores of 96% and an average IoU of 93%. The white, red, blue, and black pixels of the comparison maps represent true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN), respectively (cf. the section “Data Availability”).

**Comparisons**

Table 1 compares DeepQuantized-Net, Res-U-Net, ABF+SegNet, and DCN numerically using F1 score, IoU, and processing time (T) based on time step (tstep). The three metrics are defined as follows:

\[
F1 = \frac{2TP}{2TP + FN + FP}; \quad IoU = \frac{TP}{TP + FN + FP}; \quad T = \frac{t_{\text{step}} \times 150}{60} \tag{5}
\]

According to Table 1, significant accuracy improvements can be observed for the DeepQuantized-Net over the Res-U-Net and ABF+SegNet in terms of building extraction. Moreover, ABF+SegNet generally performed better than Res-U-Net. The function of DCN was found to be more optimized than ABF+SegNet. The proposed method achieves processing time for the building extraction of 311 minutes for the eight tested images, while DCN is 549 minutes.

Figure 5 shows the accuracy assessment of building segmentation for DeepQuantized-Net, Res-U-Net, ABF+SegNet, and DCN. According to Figure 5, the proposed method significantly improved the building segmentation results compared with using Res-U-Net, ABF+SegNet, and gated recurrent residual neural network. When the proposed method is included, the mean F1 scores and the average IoU scores could increase by approximately 7.0% and 7.2% compared to other methods, respectively.

We selected a new unsupervised method to evaluate the building extraction ability of the trained network, called multi-scale filtering building index (MFBI) (Bi et al. 2019). MFBI is a multi-scale method for building extraction to avoid complex morphological operations and use basic average filters instead. Table 2 shows the parameter settings of MFBI. In this study, due to the normalized difference vegetation index used for MFBI, it is a robust classifier for some local areas. More precisely, the visualization of segmentation results of a test image for the proposed method and MFBI are shown in Figure 6.

The function of DeepQuantized-Net was found to be more optimized than MFBI. The proposed method achieves an F1 score and the IoU score for the building extraction of 89.1% and 80.3%, while MFBI is 78.6% and 64.7%, respectively. According to Figure 6, the proposed method is effective in building segmentation capability. Although the previous method is efficient for building segmentation, it’s still not a reliable performance due to the lack of training scenarios.

In brief, four state-of-the-art deep learning and statistical methods with different backbones, including Res-U-Net, ABF+SegNet, DCN, and MFBI were used for comparisons. All networks were trained in this study from the beginning, using the same training set applied for the training of the proposed method. The benchmarking results reported in Table 1 demonstrate the complexity of the task. Comparison results illustrated that the proposed model has a competitive performance in comparison with other models.

**Ablation Experiments**

To investigate the behavior of the DeepQuantized-Net, some ablation studies are as follows. Results are shown in Table 3.

**Figure 5.** Accuracy assessment of building segmentation for the proposed method, residual network (Res-U-Net), adaptive bilateral filter + segment-based neural network (ABF+SegNet), and deformable convolutional networks (DCN). For interpreting the references to color in this figure legend, the reader is referred to the section “Data Availability.”

**Figure 6.** Comparison of building segmentation results using DeepQuantized-Net and multi-scale filtering building index. For interpreting the references to color in this figure legend, the reader is referred to the section “Data Availability.”

**Table 2.** Parameter settings of the multi-scale filtering building index.

<table>
<thead>
<tr>
<th>Threshold</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feature map</td>
<td>0.35</td>
</tr>
<tr>
<td>Normalized difference</td>
<td>0.1</td>
</tr>
<tr>
<td>vegetation index</td>
<td></td>
</tr>
<tr>
<td>Digital surface model</td>
<td>Nonbuilding: gray value &lt; 100</td>
</tr>
</tbody>
</table>

**Different Data Modalities**

Although the DeepQuantized-Net delivers good results, the building extraction was tested with different modalities, such as RGB, depth, and RGB-D. The RGB-D image can yield a much better result than the other two modalities. The overall accuracy of RGB is about 10% lower than that of fusion of RGB and Depth.
Different Backbones
DeepQuantized-Net can be set up with varying backbones for extracting building footprints. The experiments were carried out with the inception-residual network (ResNet)-v2 and ResNet-101-Feature Pyramid Network (FPN).

The experimental results showed that the trained DeepQuantized-Net with ResNet-101-FPN and RGB-D data can reliably extract building footprints in different regions.

Discussion
Building footprint maps need to be analyzed in various challenges, including different directions (e.g., sample 1), different types of roofs (e.g., sample 2), occluded objects (e.g., sample 3), complex roofs and vegetation covers (e.g., samples 4, 5, and 7), different color roofs (e.g., sample 6), and shadow problems (e.g., sample 8). Despite the previous studies that try to extract buildings from RGB images, in this paper, we have prepared a new multi-modal data set that includes very high-resolution aerial images and depth data to train an efficient network. Moreover, we proposed a novel DL model called “DeepQuantized-Net” for automatic building segmentation from a multi-modal data set.

Figure 4 shows the predicted building footprints using the DeepQuantized-Net for different directions challenge, different types of roofs, occluded objects, complex roofs and vegetation covers, different color roofs, and shadow problems. The experimental results demonstrate that the DeepQuantized-Net provides good performance on building extraction and most buildings were identified completely.

Conclusions
In this study, a novel multi-modal deep learning architecture based on superpixel-wise convolutional layers is proposed to improve automatic building segmentation with remote sensing RGB-D images. DeepQuantized-Net uses the superpixel-wise convolutional layers to extract building footprints from RGB-D images. The experimental results using the IND data set demonstrate the more reasonable accuracy and efficient computational cost achievement of the proposed method than the automatic building extraction performance of some other existing state-of-the-art DL methods. In the future, we will research more real-time building segmentation methods in the scenario of a natural disaster.

Data Availability
1. The data set generated during the current study is available at https://figshare.com/articles/dataset/IND_v2/14430766.
2. All color and grayscale images used to support this study are available at https://figshare.com/s/f280151903812c14307e.

Supplementary Information
Supplementary information for this article is available at https://figshare.com/s/40ea60d388f55097c5.
Figure S1. A map of Indiana cities that includes city boundaries. (Image courtesy of Indiana Spatial Data Portal.)
Figure S2. Building footprint correction and refinement for IND data set.
Figure S3. Building segmentation results using DeepQuantized-Net in the IND data set.
Figure S4. Data acquisition and data annotation technique.

Acknowledgments
We thank Qi Bi (Wuhan University, China) for providing the MFRI results.

Table 3. Ablation study on using different data modalities and backbones. The italic values denote the best result achieved by methods.

<table>
<thead>
<tr>
<th>Test</th>
<th>Modality</th>
<th>Backbone</th>
<th>OA (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>RGB</td>
<td>Inception-ResNet-v2</td>
<td>90.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ResNet-101-FPN</td>
<td>91.4</td>
</tr>
<tr>
<td></td>
<td>Depth</td>
<td>Inception-ResNet-v2</td>
<td>92.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ResNet-101-FPN</td>
<td>94.1</td>
</tr>
<tr>
<td></td>
<td>RGB-D</td>
<td>Inception-ResNet-v2</td>
<td>98.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ResNet-101-FPN</td>
<td>98.9</td>
</tr>
<tr>
<td>2</td>
<td>RGB</td>
<td>Inception-ResNet-v2</td>
<td>83.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ResNet-101-FPN</td>
<td>83.8</td>
</tr>
<tr>
<td></td>
<td>Depth</td>
<td>Inception-ResNet-v2</td>
<td>90.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ResNet-101-FPN</td>
<td>91.1</td>
</tr>
<tr>
<td></td>
<td>RGB-D</td>
<td>Inception-ResNet-v2</td>
<td>95.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ResNet-101-FPN</td>
<td>96.2</td>
</tr>
<tr>
<td>3</td>
<td>RGB</td>
<td>Inception-ResNet-v2</td>
<td>89.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ResNet-101-FPN</td>
<td>90.0</td>
</tr>
<tr>
<td></td>
<td>Depth</td>
<td>Inception-ResNet-v2</td>
<td>92.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ResNet-101-FPN</td>
<td>94.6</td>
</tr>
<tr>
<td></td>
<td>RGB-D</td>
<td>Inception-ResNet-v2</td>
<td>97.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ResNet-101-FPN</td>
<td>98.4</td>
</tr>
<tr>
<td>4</td>
<td>RGB</td>
<td>Inception-ResNet-v2</td>
<td>90.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ResNet-101-FPN</td>
<td>90.9</td>
</tr>
<tr>
<td></td>
<td>Depth</td>
<td>Inception-ResNet-v2</td>
<td>94.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ResNet-101-FPN</td>
<td>95.6</td>
</tr>
<tr>
<td></td>
<td>RGB-D</td>
<td>Inception-ResNet-v2</td>
<td>97.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ResNet-101-FPN</td>
<td>98.3</td>
</tr>
<tr>
<td>5</td>
<td>RGB</td>
<td>Inception-ResNet-v2</td>
<td>91.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ResNet-101-FPN</td>
<td>92.7</td>
</tr>
<tr>
<td></td>
<td>Depth</td>
<td>Inception-ResNet-v2</td>
<td>95.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ResNet-101-FPN</td>
<td>94.7</td>
</tr>
<tr>
<td></td>
<td>RGB-D</td>
<td>Inception-ResNet-v2</td>
<td>98.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ResNet-101-FPN</td>
<td>99.1</td>
</tr>
<tr>
<td>6</td>
<td>RGB</td>
<td>Inception-ResNet-v2</td>
<td>88.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ResNet-101-FPN</td>
<td>89.7</td>
</tr>
<tr>
<td></td>
<td>Depth</td>
<td>Inception-ResNet-v2</td>
<td>91.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ResNet-101-FPN</td>
<td>92.4</td>
</tr>
<tr>
<td></td>
<td>RGB-D</td>
<td>Inception-ResNet-v2</td>
<td>99.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ResNet-101-FPN</td>
<td>99.0</td>
</tr>
<tr>
<td>7</td>
<td>RGB</td>
<td>Inception-ResNet-v2</td>
<td>90.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ResNet-101-FPN</td>
<td>90.9</td>
</tr>
<tr>
<td></td>
<td>Depth</td>
<td>Inception-ResNet-v2</td>
<td>93.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ResNet-101-FPN</td>
<td>94.2</td>
</tr>
<tr>
<td></td>
<td>RGB-D</td>
<td>Inception-ResNet-v2</td>
<td>98.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ResNet-101-FPN</td>
<td>98.9</td>
</tr>
<tr>
<td>8</td>
<td>RGB</td>
<td>Inception-ResNet-v2</td>
<td>94.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ResNet-101-FPN</td>
<td>94.6</td>
</tr>
<tr>
<td></td>
<td>Depth</td>
<td>Inception-ResNet-v2</td>
<td>95.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ResNet-101-FPN</td>
<td>95.1</td>
</tr>
<tr>
<td></td>
<td>RGB-D</td>
<td>Inception-ResNet-v2</td>
<td>98.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ResNet-101-FPN</td>
<td>99.0</td>
</tr>
</tbody>
</table>

OA = overall accuracy; RGB = red, blue, green; RGB-D = red, blue, green-depth; Inception-ResNet-v2 = inception residual network v2; ResNet-101-FPN = residual network-101-Feature Pyramid Network.
References


Feature-Point Matching for Aerial and Ground Images by Exploiting Line Segment-Based Local-Region Constraints

Min Chen, Tong Fang, Qing Zhu, Xuming Ge, Zhanhao Zhang, and Xin Zhang

Abstract
In this study, we propose a feature-point matching method that is robust to viewpoint, scale, and illumination changes between aerial and ground images, to improve matching performance. First, a 3D rendering strategy is adopted to synthesize ground-view images from the 3D mesh model reconstructed from aerial images and overcome the global geometric distortion between aerial and ground images. We do not directly match feature points between the synthesized and ground images, but extract line-segment correspondences by designing a line-segment matching method that can adapt to the local geometric deformation, holes, and blurred textures on the synthesized image. Then, on the basis of the line-segment matches, local-region correspondences are constructed, and local regions on the synthesized image are propagated back to the original aerial images. Lastly, feature-point matching is performed between the aerial and ground images with the constraints of the local-region correspondences. Experimental results demonstrate that the proposed method can obtain more correct matches and higher matching precision than state-of-the-art methods. Specifically, the proposed method increases the average number of correct matches and average matching precision of the second-best method by more than five times and 40%, respectively.

Introduction
3D models are fundamental spatial information for smart city construction, and they play an important role in many fields, such as urban planning, monitoring, and management. Aerial imagery is one of the most commonly used data sets for generating city-scale 3D mesh models, because it can efficiently collect information for large-scale scenes (B. Wu et al. 2018). However, the 3D mesh models produced from aerial images may show considerable local deformations, holes, and blurred textures on the facades of the reconstructed buildings, due to occlusion and large camera tilt angles. To address these problems, combining aerial imagery with ground imagery and using the complementary advantages of two types of data sets to reconstruct complete and accurate 3D mesh models has become a research hot spot in recent years (Jende et al. 2018; B. Wu et al. 2018; Gao et al. 2019; Zhu et al. 2020).

Matching aerial and ground images to obtain enough and accurate tie points is a fundamental issue in the integration of aerial and ground images. Matching performance has an important influence on 3D reconstruction. Unfortunately, satisfactory matching results are difficult to achieve due to considerable variations in viewpoint, scale, and illumination caused by the differences between the aerial and ground platforms.

Traditional methods to extract local invariant features for matching are not robust enough for the considerable viewpoint and scale changes between aerial and ground images (Zhu et al. 2020). Some current methods try to improve the matching strategy and thus the performance of aerial–ground image matching, mainly including image correction-based methods (Gao et al. 2018; B. Wu et al. 2018) and 3D rendering-based methods (Zhu et al. 2020). These methods can eliminate or alleviate the global geometric deformation between images and improve matching performance, but some problems remain:

1. Image correction-based methods have high time complexity, because they correct each pair of images and perform feature extraction and matching on the entire image. In addition, the overlapping region between the corrected images may be too narrow in some cases to generate enough reliable correspondences.

2. For 3D rendering-based methods, the mesh models used to synthesize ground-view images often have considerable local deformations, holes, and blurred textures; thus, the synthesized images also have these problems. As the quality of synthesized images decreases, the improvement of matching performance becomes limited.

In this study, we present a robust feature-point matching method for aerial and ground images by exploiting line segment-based local-region constraints. In the proposed method, line segments are detected and matched to bridge the gap between the two types of images. Then corresponding local regions are constructed on the basis of the line-segment matches to constrain the matching of feature points between aerial and ground images. Considering the geometric deformations between aerial and ground images, we do not attempt to match line segments directly on the original aerial and ground images to construct local-region correspondences. Instead, line-segment detection and matching are performed between the ground image and the synthesized ground-view image that is produced through the 3D rendering-based strategy (Zhu et al. 2018).
Local-region correspondences are constructed on the basis of the matched line segments and propagated back to the original images. On the original aerial and ground images, all local regions are normalized to achieve affine invariance and enhanced by using the Wallis filter (Wallis 1976) to improve robustness to illumination variation. Afterward, the popular scale-invariant feature transform (SIFT; Lowe 2004) is adopted to find feature-point matches from the enhanced local regions. Lastly, a local–global constraint-based postprocessing step is designed to eliminate outliers. The proposed method performs well for aerial and ground images with considerable differences in viewpoint, scale, and illumination. Our main contributions are threefold.

First, we propose a feature-point matching framework that is robust to variances in viewpoint, scale, and illumination between aerial and ground images. The proposed method performs geometric rectification, feature detection, and matching only on the corresponding local image regions, thus avoiding the influence of nonoverlapping image regions on the matching procedure. In addition, feature-point detection and matching are not performed on the synthesized image, thus overcoming the problem of the 3D rendering-based method (Zhu et al. 2020)—that is, limited matching performance due to local geometric deformations, holes, and blurred textures on the synthesized image.

Second, a viewpoint-robust method is designed to obtain local-region correspondences to constrain feature-point matching. In contrast to traditional affine-invariant region detectors (Mikolajczyk et al. 2005) that try to extract affine-invariant regions directly, this method matches line segments first and then constructs local-region correspondences on the basis of adjacent line-segment matches. Line-segment detection and matching are conducted between the synthesized image and the ground image, thus avoiding the difficulty in matching aerial and ground images with variations in viewpoint and scale. In addition, considering local geometric deformations, holes, and blurred textures on the synthesized image, we design a line-segment matching method that relies only on loose geometric relationships.

Third, in outlier elimination, we propose a local–global constraint-based method to deal with the problems wherein the epipolar line-based method cannot eliminate outliers near the epipolar line and the global homography-based method may wrongly eliminate many correct matches. In the proposed method, a local homography-based random sample consensus (RANSAC; Fischler and Bolles 1981) algorithm is performed on the point matches in each pair of corresponding local regions. Then all the point matches on the entire image that passed the local homography-based constraint are merged and entered into an epipolar line-based RANSAC framework to eliminate additional outliers. Before local homography-based elimination, a step of local-region merging is designed to merge adjacent coplanar local regions; such merging contributes to improving the reliability of the local homography-based elimination because it reduces the possibility of performing RANSAC on a pair of small local regions that cover only a few point matches.

The remainder of this article is organized as follows: Related work is discussed in the next section, followed by details of the proposed method. Then we demonstrate and analyze the experimental results using diverse publicly available data sets. Finally, we conclude the study and provide insight into future work.

Related Work
Feature-point matching methods can be roughly divided into area-based and feature-based methods (Chen et al. 2017, 2018). Area-based methods work well on downward-looking aerial images because no considerable geometric differences—such as scale, rotation, and perspective distortions—exist between the images, benefiting from the small pitch and roll angles and the relatively stable flight height of the platform (Gruen 2012). However, the performance of area-based methods drops dramatically with increasing geometric difference between images (Zhu et al. 2020). This problem can be overcome in feature-based methods by constructing feature descriptors robust to different image variations. Therefore, for images with viewpoint and scale differences, feature-based methods are commonly used. Among them, some methods try to extract local invariant features directly for matching (Mikolajczyk and Schmid 2004; Mikolajczyk et al. 2005), whereas others first design a preprocessing step to alleviate the geometric deformations between images (Gao et al. 2018; B. Wu et al. 2018; Zhu et al. 2020) and then use traditional feature-matching methods to find point correspondences. In this study, these two types of feature-based methods are called, respectively, direct and indirect matching methods.

Direct Matching Methods
The SIFT algorithm (Lowe 2004) is one of the most widely used direct matching methods, due to its scale and rotation invariance. However, its performance decreases as the viewpoint changes, because it is not invariant to image viewpoint variation. To overcome this problem, some affine-invariant region detectors have been proposed, such as the maximally stable extremal region (Matas et al. 2004) and the intensity extreme-based region (Thytylaars and Van Gool 2004). However, the region detectors can obtain a few features and achieve a low feature-repetition rate in practical applications, because they are usually sensitive to image noise. They are still not robust enough to variations in image viewpoint and scale to deal with aerial and ground images.

With the development of artificial intelligence, deep learning-based frameworks have been introduced into feature-point matching. Among them, some methods simultaneously incorporate feature description and similarity measurement into a unified deep neural network and learn feature descriptors and similarity measures by maximizing negative sample distance (noncorresponding features) and minimizing positive sample distance (corresponding features; Han et al. 2015; Zagoruyko and Komodakis 2015; Kumar et al. 2016; S. Wang et al. 2018; He et al. 2019). Some other methods only learn feature descriptors through a deep neural network. The latter methods have attracted attention because of their great flexibility in use while ensuring matching performance (Simoes-Serra et al. 2015; Mishchuk et al. 2017; Tian et al. 2017). Among the deep learning-based methods, SuperPoint (DeTone et al. 2018), GeoGlue (Luo et al. 2018), and ContextDesc (Luo et al. 2019) have shown excellent performance on images with geometric deformations. Furthermore, an attentional graph neural network-based similarity measurement method called SuperGlue (Sarlin et al. 2020) has been proposed. Combined with SuperPoint, it achieved impressive matching performance on images with considerable scale and viewpoint changes. However, we did not obtain the expected results when we applied this method on aerial and ground images, due to the limitations of the robustness of the learned features to scale and viewpoint changes (see details under “Experimental Results and Analysis”).

Indirect Matching Methods
The similarity of corresponding image regions becomes remarkably reduced as the difference in image viewpoint increases, making it difficult to extract invariant features directly from images with large viewpoint changes. Therefore, many studies do not attempt to extract invariant features
directly for matching, but instead focus on how to improve the matching strategy to enhance performance. This has gradually become the mainstream idea for matching images with large viewpoint changes (Zhu et al. 2020).

For images without any prior information, researchers propose simulating an image affine or projective space first and then performing feature matching in the simulated image space (Morel and Yu 2009; Cai et al. 2013). This type of method can obtain more matches than SIFT on images with large viewpoint changes. However, the time efficiency of these methods is much lower than that of SIFT, because they perform feature extraction and exhaustive matching in the entire simulated space, resulting in a multiple increase in the number of extracted features and of feature similarity comparisons. Based on the idea of image simulation, some iterative matching frameworks have been proposed to improve the time efficiency (Yu et al. 2012; Chen et al. 2013a). However, studies have found that such methods do not perform well on aerial and ground images (Shan et al. 2014).

To overcome the scale change between aerial and ground images, we can first construct a scale space and use methods such as bag of features (Philbin et al. 2007) to describe each image layer in the scale space for aerial and ground images, respectively; then find the corresponding image layers with similar scales between the two scale spaces; and finally perform feature-point matching on corresponding image layers, which can effectively improve the precision and time efficiency of feature matching (Zhou et al. 2017). However, given the large viewpoint changes between aerial and ground images, the reliability of the image-layer correspondences estimated using the SIFT feature-based bag of features is difficult to guarantee.

In the field of photogrammetry, prior information, such as data from a position and orientation system, is usually used as auxiliary information to perform geometric correction and eliminate or alleviate the global geometric deformation between images. Afterward, traditional feature-matching methods, such as SIFT, can be used to find point matches from the corrected images. We call these methods geometric correction-based methods, which can be further subdivided into view-independent and view-dependent correction methods.

View-independent methods usually find a common base plane and project all images onto this base plane to alleviate the geometric distortion between them (Hu et al. 2015; Jiang and Jiang 2017). However, finding a suitable base plane to project all the images onto is not always possible. To overcome this problem, some view-dependent methods have been proposed that individually correct each pair of images. Gao et al. (2018) performed sparse reconstruction for aerial and ground images separately and then synthesized the ground images into the aerial views on the basis of the reconstructed mesh models and the calculated exterior orientation elements to alleviate the geometric distortions. Considering that the overlapping areas that are useful for image matching between a pair of aerial and ground images in urban areas mainly cover the facades of buildings, B. Wu et al. (2018) corrected aerial and ground images to a building facade-based base plane to obtain images that are easy to match by exploiting the exterior orientation elements of the aerial and ground images and the dense point cloud from the aerial images. These view-dependent methods show excellent performance in the matching of aerial and ground images. However, time efficiency is relatively lower than with the view-independent methods. When an image is paired with multiple different images, it is corrected multiple times, and then feature extraction is performed separately for each corrected image. In addition, given the narrow overlapping range of aerial and ground images, view-independent and view-dependent methods are susceptible to interference from numerous features in nonoverlapping areas in the feature-matching step; such susceptibility is not conducive to improving the matching precision.

To solve these problems of geometric correction-based methods, Zhu et al. (2020) proposed a 3D rendering-based method called MeshMatch. This method performs dense reconstruction on aerial images to generate a 3D mesh model and sparse reconstruction on ground images to obtain the exterior orientation parameters; then it synthesizes ground-view images by rendering the aerial image-based mesh model at the position of the ground cameras. SIFT feature-point matching is performed on the ground and synthesized ground-view images. Matching performance can be remarkably improved because the global geometric deformations and occlusions are eliminated between the ground and synthesized ground-view images. In addition, given that one synthesized image corresponds to multi-view aerial images, the matched points on the synthesized image can be propagated to multiple aerial images to obtain the matching results of multiple pairs of aerial and ground images after feature matching on a pair of synthesized and ground images, thus effectively improving time efficiency. However, given that synthesized images often show considerable local geometric deformations, holes, and blurred textures due to the quality of the 3D mesh model, performing SIFT or other grayscale-dependent feature-matching methods directly on the synthesized images is not suitable.

In view of the problems of the 3D rendering-based method, we first design a line-segment matching method that relies only on geometric relations to obtain line-segment matches between the synthesized image and the ground image. Then local-region correspondences are constructed on the basis of line-segment matches and propagated back to the aerial and ground images. Afterward, feature-point extraction and matching are conducted under the constraints of the local-region correspondences. The proposed method is robust to variations in viewpoint, scale, and illumination between aerial and ground images, while avoiding the negative influence of geometric and textural defects of the synthesized image on the matching performance.

**Methodology**

The flowchart of the proposed method is shown in Figure 1. First, ground-view images are synthesized from the aerial image-based 3D mesh model using the method of Zhu et al. (2020). Considering that geometric and textural defects in the synthesized image make it difficult for gray information-based line-segment matching methods to achieve good matching

**Figure 1. Flowchart of the proposed method.**
performance, we propose a line-segment matching method that relies only on geometric relationships between line segments to generate line-segment matches and construct local-region correspondences. Afterward, the local regions on the synthesized image are propagated back to the aerial images to guide the feature-point matching between aerial and ground images. Finally, a method that integrates local and global constraints is designed to eliminate outliers.

**Construction of Local-Region Correspondences for Synthesized and Ground Images**

The enlarged subimage of the synthesized image shown in Figure 2 reveals that although directly performing feature-point matching is not suitable, due to the considerable local geometric and textural defects, the structure of many targets on the image remains prominent. Therefore, we can try to extract the local regions covered by corresponding local structures on the synthesized and ground images and propagate them to the aerial and ground images to constrain feature-point matching. In this study, we detect line segments and propose a robust method to obtain line-segment matches to construct local-region correspondences.

![Figure 2. Example of a synthesized image to show the local geometric and textural defects and the target structure.](image)

**Line-Segment Detection**

Line-segment detection methods include hand-designed and deep learning-based methods. The existing deep learning-based methods rely on training samples, and their generalization ability is limited. Therefore, hand-designed methods are more reliable in practical applications. Among hand-designed methods, LSD (von Gioi et al. 2010) and EDLines (Akinlar and Topal 2011) are two of the most widely used (K. Li et al. 2016a). However, when we use these two methods to extract line segments from a synthesized image, the following problems will occur: first, many long line segments are fragmented into short line segments (see the line segments in the white rectangular boxes in Figure 3A and 3B). Even if some of the fragmented short line segments can be matched successfully, the short lines can construct only small local regions to cover a small number of feature points in the feature-point matching procedure. In addition, small local regions are not conducive to generating distinctive feature descriptors to improve the performance of feature-point matching. Second, line segments in weakly textured areas are difficult to successfully extract using the LSD and EDLines detectors (see the areas in the white ovals in Figure 3A and 3B). Based on these considerations, we do not directly use the existing line-segment detectors, but instead we design a line-segment extraction method that is more suitable for the synthesized image.

In this study, the method proposed by Chen et al. (2013b) is improved to detect line segments for synthesized and ground images. Edge pixels on the synthesized and ground images are extracted first using the Canny operator (Canny 1986), and then edge splitting and fitting are performed to generate line segments. Given the local deformations, holes, and blurred textures on the synthesized images, line segments with a high repetition rate between the synthesized and ground images are difficult to obtain using the same edge-splitting threshold. For this reason, in edge splitting we set a loose and a strict threshold for the synthesized and ground images, respectively (7 and 5 pixels). In addition, a step of line-segment merging is performed to merge adjacent collinear line segments.

**Geometric Relationship-Based Line-Segment Matching**

Many line-segment matching methods have been proposed in the literature. Some methods adopt the geometric and grayscale information around individual line segments to construct descriptors and measure the similarity of these line segments (Schmid and Zisserman 1997; Bay et al. 2005; Z. Wang et al. 2009). These methods require that the corresponding line segments have sufficient overlap and consistent illumination or color to extract similar descriptors. Two or more line segments in an image are usually used to form a line group, and each line group is regarded as one feature to construct constraints in the matching procedure (L. Wang et al. 2009; Kim and Lee 2012; Ok et al. 2012; Al-Shahri and Yilmaz 2014; Yammine et al. 2014; López et al. 2015; K. Li et al. 2016b; K. Li and Yao 2017). Such methods can generally achieve better matching performance than individual line segment-based methods, because more constraints for disambiguation can be constructed through line groups. However, as the number of line segments in each group increases, the complexity of constructing geometric constraints will increase significantly.

In addition, geometric invariants can be constructed to find line-segment correspondences on the basis of the geometric relationship between point correspondences and line segments (Fan et al. 2012; Chen and Shao 2013; Ramalingam et al. 2015; Sun et al. 2015; Jia et al. 2018; J. Wang et al. 2021). However, substantial point correspondences should be known as prior knowledge to construct geometric invariants. Therefore, these methods are limited in practical applications.

Considering the local deformations, holes, and blurred textures on synthesized images, using gray information-based matching methods is inappropriate. Besides, since line-segment matching is only a preprocessing step of the feature-point matching method proposed in this article, the time efficiency of line-segment matching is a problem that needs to be considered. The existing line-segment matching methods are not suitable as a preprocessing step of the proposed feature-point matching...
matching method. Thus, we propose an efficient line-segment matching method that relies only on the geometric relations between line segments, based on the assumption that no obvious global geometric distortion exists between the synthesized and ground images. First, line segments are grouped in accordance with the position relationship between them (L. Wang et al. 2009). Afterward, line groups are compared by measuring their similarity. With the matched line groups, we can find the corresponding line segments on the basis of the relative position relationship of line segments in the groups.

**Similarity measure of line groups:** The similarity of two groups of line segments is equal to the sum of the similarities of every two line-segment pairs in the two line groups. Given that the corresponding line segments in the two line groups are unknown before the two groups are matched, we calculate the similarity of the two groups under each different case of line-segment correspondence and finally select the maximum similarity as the similarity of the two line groups.

In accordance with the line-segment grouping rule (L. Wang et al. 2009), the local area between two line segments in a line-segment group can be approximated as a flat area, and the transformation between the local areas of two corresponding line-segment pairs can be approximated as an affine transformation. Therefore, two affine invariants based on line-segment length ratios (EC/CD and EA/AB in Figure 4) are used in measuring the similarity of two line-segment pairs. In addition, given that no obvious global geometric difference exists between the synthesized and ground images, two corresponding line segments should be approximately parallel, and the distance between the intersection points of two corresponding line-segment pairs should be close to 0. Based on these assumptions, we calculate the similarity of two line-segment pairs \( LP_{ll} \) and \( LP'_{ll} \) according to

\[
\text{Sim}(LP_{ll}, LP'_{ll}) = \begin{cases} 
  d_{s1} + d_{s2} + d_{s} + d_{s}, & \text{if } \Gamma = \text{true}, \\
  0, & \text{otherwise} 
\end{cases}
\]

where \( d_{s1} = 1 - \frac{a_{ll}l_{s}}{T_{s}}, \)
\( d_{s2} = 1 - \frac{a_{ll}l'_{s}}{T_{s}}, \)
\( d_{s} = 1 - \frac{|a_{ll} - a_{ll}l|}{T_{s}}, \)
\( d_{s} = 1 - \frac{\min(|r_{s}l - r_{s}l' |, |r_{s}l' - r_{s}l|)}{T_{s}}, \)
\( \Gamma = \{d_{s1}, d_{s2}, d_{s}, d_{s}\} \geq 0, \quad x_{s} - x_{s} < T_{s} \)

where \( l_{s} \) and \( l'_{s} \) represent the line segments forming a line pair on the synthesized image; \( l_{s} \) and \( l'_{s} \) are the line segments in the line pair to be compared on the ground image; \( a_{ll} \) and \( a_{ll}l' \) are the rotation angles between line segments \( l_{s} \) and \( l'_{s} \) and line segments \( l_{s} \) and \( l'_{s} \), respectively; \( r_{s}l \) and \( r_{s}l' \) are the angles between line segments \( l_{s} \) and \( l'_{s} \) and line segments \( l_{s} \) and \( l'_{s} \), respectively; \( r_{s}l = EC/CD \) and \( r_{s}l' = EA/AB \) are the length ratios as shown in Figure 4; \( r_{s}l \) and \( r_{s}l' \) are the ratios corresponding to \( r_{s}l \) and \( r_{s}l' \) in the compared line pair; \( x_{s} \) and \( x_{s} \) are the image coordinates of the intersection of the two line-segment pairs; and \( T_{s}, T_{s}, T_{s} \) are thresholds.

**Fast and reliable line-segment correspondence searching:** For each line group on the synthesized image, we find the line group with the greatest similarity on the ground image. If the similarity value is greater than a threshold \( T_{s} \), then the two line groups are considered matched. In this procedure, the following strategies are used to improve matching efficiency and reliability.

For any line group \( LG = \{l_{1}, l'_{1}, \cdots, l'_{k}\} \) on the synthesized image, where \( l_{k} \) is the central line segment in the group and the other \( k \) line segments are grouped with \( l_{k} \) to form \( LG \), we calculate its minimum bounding rectangle and map the rectangle to the ground image. Considering that the global geometric transformation between the synthesized and ground images is small, we map the rectangle to the same position on the ground image as on the synthesized image. Then we find the line segments on the ground image that have at least one pixel in the rectangle. Line groups are formed on the basis of these line segments to be the candidate matches of the line group \( LG \), on the synthesized image. The following process is performed only on the line groups in the set of candidate matches.

To compute the similarity between the line group \( LG \) and each of the candidate matches \( LG' = \{l_{1}, l'_{1}, \cdots, l'_{k}\} \) efficiently, we calculate the similarity between two line groups only when there are more than two pairs of line segments between \( \{l_{1}, l'_{1}, \cdots, l_{k}\} \) and \( \{l_{1}, l'_{1}, \cdots, l_{k}\} \) whose direction deviation is less than \( T_{s} \); otherwise, we directly set the similarity to 0.

In addition, for \( LG \), we pair the center line segment \( l'_{k} \) with other line segments to form \( k \) line pairs \( \{LP_{kk}, LP_{kk}, \cdots, LP_{kk}\} \). We also obtain \( k \) line pairs \( \{LP_{kk}, LP_{kk}, \cdots, LP_{kk}\} \) for the candidate match \( LG' \). Only when the maximum line-pair similarity

\[
\max \left\{ S(LP_{kk}', LP_{kk}) \right\} > 0
\]

we calculate the similarity of all line pairs (including line pairs formed with two noncentral line segments) to obtain the

![Figure 4. Illustration of parameters in the measurement of line-pair similarity: (A, B) two line pairs to be compared.](image)
similarity of the two line groups; otherwise, we directly set the similarity of the two line groups to 0.

Construction of Local-Region Correspondences

Due to the severe local geometric distortions on the synthesized image, the line segments on the synthesized image can only roughly represent the target contour. The position and orientation deviation of the line segments are very large. In this case, the deviation of the intersection of a pair of line segments is often very large; especially when the angle between the two line segments is small, the deviation between their intersection and the intersection of the two corresponding line segments is even more unacceptable. In addition, even if the position and orientation of line segments are very accurate, two pairs of matched line segments may intersect to generate false point matches because the adjacent line segments may be not coplanar. Therefore, in this paper we do not directly obtain point matches by intersecting matched line segments but construct local regions with matched line segments to constrain the feature-point matching, and then obtain accurate point matches.

The line-segment pairing method of K. Li et al. (2016b) is used to generate line pairs based on the matched line segments on the synthesized image. Corresponding line-segment pairs on the ground image can be constructed in accordance with the corresponding relationship between the matched line segments on the ground image and the synthesized image. Line pairs may be of either V or T type. A parallelogram region is formed for each line pair, as shown in Figure 5. Therefore, many corresponding parallelogram regions can be obtained between the synthesized and ground images on the basis of the line-pair correspondences.

Given that line-segment detection is affected by noise and occlusion, the endpoints of corresponding line segments are not correspondences, which leads to a reduction in the consistency of the corresponding parallelogram regions constructed on the basis of the line-pair correspondences. To overcome this problem, we first estimate the fundamental matrix between the synthesized image and the ground image through the RANSAC algorithm, based on the intersections of corresponding line pairs. Then, with the fundamental matrix, a local homography is estimated for each pair of corresponding line pairs (K. Li and Yao 2017). Afterward, the vertices of the parallelogram on the synthesized image are mapped onto the ground image on the basis of the local homography to generate a pair of parallelogram regions with corresponding vertices. In this manner, we can obtain many local-region correspondences with similar image content between the synthesized and ground images.

Local Region-Based Feature-Point Matching for Aerial and Ground Images

Given that the synthesized image is rendered from the aerial image-based mesh model, the local regions on the synthesized image can be propagated back to the aerial images to generate corresponding local regions between aerial and ground images. Then geometric normalization and grayscale enhancement can be performed on the local regions to achieve affine and illumination invariance.

Propagation of Local Regions to Aerial Images

In accordance with the coordinates and depth information of the vertices, the local regions on the synthesized image can be propagated to the aerial images. The following problems may occur during the propagation: Some vertices of the local regions on the synthesized image may fall on nearby targets with discontinuous parallax, due to the inaccurate endpoints of the detected line segments (Figure 6A), resulting in a large deviation between the local region on the aerial image and the corresponding local region on the ground image. Some vertices of the local regions on the synthesized image may fall in an area (e.g., holes) without a depth value (Figure 6B), resulting in failure of the propagation. And some local regions on the synthesized image may be visible but the corresponding regions on some aerial images be invisible (Figure 6C), due to viewpoint variation and occlusion.
To deal with these problems, for each of the two line segments $l^1$ and $l^2$ forming a local region on the synthesized image, we find seven points on the line segment: the two endpoints and five points dividing the line segment into six segments of equal length. Among the seven points, if a point $p$ without a depth value exists, then we find the closest pixel to $p$ with a depth value and assign that depth value to $p$. On the basis of the depth values, the seven points are mapped to the aerial image. On the aerial image, the RANSAC algorithm is used to find the inliers from the seven points that best fit a straight line. Then the corresponding points of these inliers are found from the seven points on the synthesized image. In this manner, for $l^1$ and $l^2$, two inlier point sets $P_1$ and $P_2$ can be obtained on the synthesized image. We take a point each from $P_1$ and $P_2$ to combine with the intersection of $l^1$ and $l^2$; then we construct a parallelogram region according to the method shown in Figure 5 and propagate the parallelogram region to the aerial images. All points in $P_1$ and $P_2$ are used to form a parallelogram region iteratively. Among all the local regions formed from $P_1$ and $P_2$, the one with the largest area and all four vertices within the image range is regarded as the final local region of $l^1$ and $l^2$. When a local region on the synthesized image is determined, its corresponding local region on the ground image can be determined in accordance with the corresponding points on the two line segments of the corresponding line pair. In addition, for a local region propagated back to the aerial image, if its area is less than $T_o = 1/10$ (determined empirically) of the area of the local region on the synthesized image, then it is considered an invisible area with a high probability. Therefore, it is an unreliable local region and thus removed.

**Local Region-Based Feature-Point Matching**

After obtaining the corresponding local regions between the aerial image and the ground image, feature-point matching can be performed on the local regions. To improve the performance of feature-point matching, the corresponding local regions are normalized into square regions of the same size, which can eliminate the geometric deformation between the corresponding local regions caused by viewpoint change. Furthermore, in consideration of the illumination variation or shadows between aerial and ground images, the normalized local images are enhanced using the Wallis filter to improve the similarity of the local images.

The SIFT method is performed on the enhanced corresponding local regions to find point matches. Given that the geometric deformation between the corresponding local regions has been eliminated through normalization, for each feature point in the local region on the aerial image, the search for the corresponding point is performed only in a search region ($T_P$ pixels) centered on the feature point in the local region of the ground image. The point matches obtained in each pair of local regions are calculated back to the aerial and ground images to obtain the final matches.

**Outlier Elimination**

A local–global constraint-based method is proposed to eliminate outliers. Given that the local region can be approximated as a plane region, the homography-based RANSAC algorithm can be used to eliminate outliers of the matches in each pair of local regions (Hartley and Zisserman 2000). However, performing RANSAC on every pair of local regions is inefficient and unreliable, due to the numerous local regions and the small area of some local regions that can cover only a few initial point matches. To overcome this problem, a step of local-region merging is performed before the local homography-based outlier elimination.

For any two local regions $A$ and $B$, if the overlapping area of $A$ and $B$ satisfies the condition in Equation 3, then the two local regions are considered approximately coplanar. Then the point matches in $A$ and $B$ are merged, and the local homography-based RANSAC is performed on the set of merged point matches to eliminate outliers. When the point matches in all local regions are checked using the local homography-based RANSAC algorithm, all the inliers are merged and entered into an epipolar line-based RANSAC framework to eliminate additional outliers:

$$R_{\text{overlap}} / \min(R_A, R_B) > T_R$$

where $R_{\text{overlap}}$ represents the overlapping area of $A$ and $B$, $R_A$ and $R_B$ represent the respective areas of $A$ and $B$, $\min( )$ is a function to detect the minimum value, and $T_R$ is a threshold.

**Experimental Results and Analysis**

**Experimental Data Sets and Evaluation Criteria**

As shown in Figure 7, 10 pairs of publicly available aerial and ground images are used to evaluate the proposed matching method. The image pairs 1–2 and 3–4 are from the International Society for Photogrammetry and Remote Sensing benchmark data sets (collected at Centre of Dortmund and Zeeche of Zurich, respectively; Nex et al. 2015). The other six pairs of images were collected on the campus of Southwest Jiaotong University (Zhu et al. 2020). Detailed information of the experimental data sets is provided in Table 1.

In the experiments, two widely used indicators (J. Li et al., 2017)—number of correct matches (NCM) and matching precision (MP)—are adopted to evaluate the performance of the proposed method. MP is calculated as

$$\text{MP} = \frac{(\text{NCM})}{\text{NTM}}$$

where NTM is the number of total matches. We manually check the matches to obtain the NCM value.

**Threshold Setting**

Table 2 lists all the relevant parameters in the proposed method. The threshold $T_o$ is set at 0.3 according to L. Wang et al. (2009). The area size ratio $T_R = 0.1$ is determined empirically. We recommend setting a small threshold for $T_o$, because a large one may wrongly delete some local regions that are correctly propagated back to the aerial image. Although a smaller threshold may not be able to delete all the incorrectly propagated local regions, there is a step of outlier elimination after the feature-point matching, and the false point matches generated in the incorrectly propagated local regions can be eliminated in that step. We randomly select one pair of images from each of the five data sets described in Table 1 to evaluate the five other thresholds $T_o$, $T_{\alpha}$, $T_{\beta}$, $T_{\alpha'}$, and $T_{\beta'}$. Two indicators, average NCM and average MP on all test image pairs, are used in the evaluation.

**Table 2. Thresholds in the proposed method.**

<table>
<thead>
<tr>
<th>Threshold Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_o$</td>
<td>Angle difference, length-ratio difference, intersection coordinate deviation, and similarity thresholds in line-segment matching</td>
</tr>
<tr>
<td>$T_{\alpha}$</td>
<td>Region-size ratio in the propagation of local regions to aerial images</td>
</tr>
<tr>
<td>$T_{\beta}$</td>
<td>Size of the search region in feature-point matching</td>
</tr>
<tr>
<td>$T_R$</td>
<td>Overlap threshold in outlier elimination</td>
</tr>
</tbody>
</table>

When we are evaluating one parameter, the other parameters are treated as constants. The statistical results of the average NCM and average MP on all test image pairs using the proposed method with different thresholds are shown in Figure 8. In our tests, when the average NCM reaches a certain level, the average MP is considered to be a more important indicator. Based on this principle, it can be seen from the figure that when $T_o$, $T_{\alpha}$, $T_{\beta}$, and $T_R$ are set to 12°, 20, 3, 120, and
Figure 7. Experimental data sets. (a, b) Two pairs of images from the International Society of Photogrammetry and Remote Sensing benchmark data sets collected at Centre of Dortmund. (c, d) Two pairs of images from the benchmark data sets collected at Zeche of Zurich. (e–j) Six pairs of images captured from the campus of Southwest Jiaotong University.

Table 1. Detailed information of the experimental data sets.

<table>
<thead>
<tr>
<th>No.</th>
<th>Test Site</th>
<th>Sensor</th>
<th>GSD (cm)</th>
<th>Image Size (pixels)</th>
<th>Convergent Angle (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Dortmund</td>
<td>Sony Nex-7</td>
<td>Aerial: 1.10, ground: 0.53</td>
<td>4000 × 6000</td>
<td>50.8</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>61.9</td>
</tr>
<tr>
<td>3</td>
<td>Zurich</td>
<td>Sony Nex-7</td>
<td>Aerial: 0.56, ground: 0.28</td>
<td>4000 × 6000</td>
<td>40.9</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>51.5</td>
</tr>
<tr>
<td>5</td>
<td>SWJTU</td>
<td>Aerial: Sony ICLE-5100, ground: Canon EOS M6</td>
<td>Aerial: 1.69, ground: 1.06</td>
<td>4000 × 6000</td>
<td>54.6</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>61.2</td>
</tr>
<tr>
<td>7</td>
<td>SWJTU</td>
<td>Aerial: Sony ICLE-5100, ground: Canon EOS M6</td>
<td>Aerial: 1.93, ground: 1.33</td>
<td>4000 × 6000</td>
<td>59.6</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>70.2</td>
</tr>
<tr>
<td>9</td>
<td>SWJTU</td>
<td>Aerial: Sony ICLE-5100, ground: DJI spark</td>
<td>Aerial: 1.97, ground: 2.56</td>
<td>Aerial: 4000 × 6000, ground: 3040 × 4056</td>
<td>34.6</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>75.1</td>
</tr>
</tbody>
</table>

SWJTU = Southwest Jiaotong University.
0.2, respectively, the proposed method achieves good overall performance. Specifically, it can be seen from Figure 8D and 8E that the search-region size and overlap thresholds do not significantly affect the matching performance.

**Evaluation of the Proposed Line-Segment Matching Method**

Line-segment matching between the synthesized and ground images is an important step in the proposed feature-point matching method. We evaluate the proposed line-segment matching algorithm on the 10 pairs of images by comparing it with the LS (L. Wang et al. 2009), MSLD (Z. Wang et al. 2009), LJL (K. Li et al. 2016b), and N-LPI (Jia et al. 2018) algorithms in this subsection.

Figure 9 illustrates the matching performance of the compared methods on all experimental image pairs in terms of NCM and completeness. In our experiments, completeness is measured by the sum of the lengths of the line segments.
that are correctly matched. Results in the figure show that the LJL method achieves the best performance in terms of NCM and completeness. The performance of the proposed method is not as good as those of LJL or N-LPI on those two aspects. However, compared with LS (a geometric relationship-based method, as is the proposed method), the proposed method greatly increases the NCM values on all image pairs while effectively improving the completeness.

Because line-segment matching is a preprocessing step of the proposed feature-point matching method, time efficiency is a very important indicator to evaluate the compared methods. Table 3 shows the running times of all compared line-segment matching methods. It can be seen from Table 3 that the time efficiency of the LJL and N-LPI methods is very low. Even if these two methods perform better than other methods in terms of NCM and completeness, they are unacceptable in practical applications. The time efficiency of MSLD, LS, and the proposed method is significantly higher than those of LJL and N-LPI. Among these three methods, the time efficiency of the proposed method is higher than that of LS and slightly lower than that of MSLD. Considering the performance in terms of NCM, completeness, and time efficiency, the proposed method has better comprehensive performance and is more suitable as a preprocessing step of the proposed feature-point matching method.

### Evaluation of the Proposed Aerial–Ground Imagery Feature-Point Matching Method

In this subsection, we compare the proposed feature-point matching method with state-of-the-art robust feature-point matching methods. The matching comparison by Zhu et al. (2020) revealed that the MeshMatch method performs better than the deep learning-based R2D2 method (Revaud et al. 2019) and the matching methods of VisualSFM (C. Wu 2011), Agisoft Metashape (Agisoft, St. Petersburg, Russia), and Colmap (Schönberger and Frahm 2016; Schönberger et al. 2016) for aerial and ground images. Therefore, in our experiments, we compare the proposed method with MeshMatch instead of R2D2, VisualSFM, Agisoft Metashape, and Colmap. In addition, we compare the proposed method with two popular handcrafted methods—SIFT (Lowe 2004) and ASIFT (Morel and Yu 2009)—and a deep learning-based method: SuperGlue (Sarlin et al. 2020). SuperGlue achieved excellent results in several competitions at the 2020 IEEE Conference on Computer Vision and Pattern Recognition.

Tables 4 and 5 show a quantitative comparison in terms of NCM and MP between the proposed method and SIFT, ASIFT, MeshMatch, and SuperGlue on all image pairs. The statistical results for NCM in Table 4 reveal that the proposed method obtained the most correct matches on all image pairs. MeshMatch and SuperGlue achieved the second- and third-best overall results, respectively. SIFT and ASIFT performed poorly, failing to yield correct matches on almost all image pairs.

### Table 3. Running time (s) of each method on all image pairs.

<table>
<thead>
<tr>
<th>Image Pair</th>
<th>LS</th>
<th>MSLD</th>
<th>LJL</th>
<th>N-LPI</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>17.80</td>
<td>1.55</td>
<td>5161.38</td>
<td>4048.37</td>
<td>4.95</td>
</tr>
<tr>
<td>2</td>
<td>17.39</td>
<td>1.31</td>
<td>4322.02</td>
<td>2890.93</td>
<td>3.33</td>
</tr>
<tr>
<td>3</td>
<td>11.70</td>
<td>1.05</td>
<td>1330.55</td>
<td>2305.45</td>
<td>3.84</td>
</tr>
<tr>
<td>4</td>
<td>10.57</td>
<td>1.19</td>
<td>945.87</td>
<td>3362.81</td>
<td>3.44</td>
</tr>
<tr>
<td>5</td>
<td>7.11</td>
<td>1.21</td>
<td>869.09</td>
<td>2319.26</td>
<td>2.91</td>
</tr>
<tr>
<td>6</td>
<td>8.03</td>
<td>1.11</td>
<td>476.68</td>
<td>1952.00</td>
<td>2.92</td>
</tr>
<tr>
<td>7</td>
<td>8.55</td>
<td>1.88</td>
<td>1044.40</td>
<td>2543.40</td>
<td>4.92</td>
</tr>
<tr>
<td>8</td>
<td>7.61</td>
<td>1.49</td>
<td>1707.69</td>
<td>3243.94</td>
<td>3.68</td>
</tr>
<tr>
<td>9</td>
<td>27.88</td>
<td>2.48</td>
<td>9750.69</td>
<td>10516.74</td>
<td>7.28</td>
</tr>
<tr>
<td>10</td>
<td>11.60</td>
<td>2.22</td>
<td>6660.83</td>
<td>9681.99</td>
<td>5.98</td>
</tr>
</tbody>
</table>

### Table 4. Number of correct matches for each method on all image pairs.

<table>
<thead>
<tr>
<th>Image Pair</th>
<th>SIFT</th>
<th>ASIFT</th>
<th>MeshMatch</th>
<th>SuperGlue</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>244</td>
<td>116</td>
<td>2269</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>98</td>
<td>111</td>
<td>1358</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>232</td>
<td>0</td>
<td>711</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>110</td>
<td>0</td>
<td>180</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>6</td>
<td>41</td>
<td>256</td>
<td>503</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>3</td>
<td>34</td>
<td>316</td>
<td>426</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>272</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>220</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>59</td>
<td>0</td>
<td>126</td>
</tr>
<tr>
<td>10</td>
<td>32</td>
<td>0</td>
<td>32</td>
<td>65</td>
<td>169</td>
</tr>
</tbody>
</table>

Boldface indicates the best performance on an image pair.

### Table 5. Matching precision (%) for each method on all image pairs.

<table>
<thead>
<tr>
<th>Image Pair</th>
<th>SIFT</th>
<th>ASIFT</th>
<th>MeshMatch</th>
<th>SuperGlue</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.27</td>
<td>0.00</td>
<td>98.39</td>
<td>75.82</td>
<td>99.39</td>
</tr>
<tr>
<td>2</td>
<td>0.00</td>
<td>0.00</td>
<td>95.15</td>
<td>72.55</td>
<td>99.34</td>
</tr>
<tr>
<td>3</td>
<td>0.06</td>
<td>0.00</td>
<td>97.89</td>
<td>0.00</td>
<td>96.60</td>
</tr>
<tr>
<td>4</td>
<td>0.00</td>
<td>0.00</td>
<td>93.22</td>
<td>0.00</td>
<td>89.55</td>
</tr>
<tr>
<td>5</td>
<td>2.17</td>
<td>19.35</td>
<td>52.56</td>
<td>69.00</td>
<td>98.63</td>
</tr>
<tr>
<td>6</td>
<td>0.00</td>
<td>9.68</td>
<td>58.62</td>
<td>79.40</td>
<td>98.61</td>
</tr>
<tr>
<td>7</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>76.19</td>
</tr>
<tr>
<td>8</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>94.02</td>
</tr>
<tr>
<td>9</td>
<td>0.00</td>
<td>0.00</td>
<td>84.29</td>
<td>0.00</td>
<td>88.11</td>
</tr>
<tr>
<td>10</td>
<td>11.43</td>
<td>0.00</td>
<td>52.46</td>
<td>32.34</td>
<td>71.91</td>
</tr>
</tbody>
</table>

Boldface indicates the best performance on an image pair.

Specifically, compared with the second-best method (MeshMatch), the proposed method increased the average NCM by more than five times (from 85 to 623.4). Especially for image pairs 7 and 8, all the compared methods failed to obtain a correct match, but the proposed method still obtained more than 200 pairs of correct matches on each image pair. The statistical results for MP reveal that the proposed method achieved...
the highest MP on 80% of the image pairs. Although the MP of the proposed method is lower than that of MeshMatch on image pairs 3 and 4, the difference is not too large. Overall, compared with the second-best method (MeshMatch), the proposed method improves the average MP by more than 40% (from 63.26% to 91.23%). Figure 10 shows the matches obtained by the proposed method on each image pair.

The proposed method can achieve such a large improvement over MeshMatch mainly due to the following two improvements: First, feature-point matching in the proposed method is performed on the original aerial image rather than on the synthesized image. In this manner, the negative influence on the matching caused by local geometric deformation, holes, and blurred textures in the synthesized image can be avoided. The local geometric deformations and blurred textures result in low localization accuracy of the features and low reliability of the calculated feature descriptors on the synthesized image. MeshMatch uses the matches on the synthesized image as the initial matches and performs template matching on the original aerial image to improve localization accuracy. However, when the initial match has a large deviation from the real corresponding point, this strategy hardly achieves the expected performance.

Second, feature-point matching in the proposed method is performed under the constraints of local-region correspondences, which can greatly reduce the search range of candidate point matches and reduce the interference of numerous noncorresponding points and the interference of repeated textures on the matching process.

SuperGlue showed good performance on image pairs 2, 5, 6, and 10 but failed to obtain correct matches on image pairs 3, 4, 7, 8, and 9. The performance of this method on different images is not stable enough, thus limiting its use in practical applications. Further research is necessary to improve its adaptability to different scenarios. In addition, SuperGlue failed on image pairs 3, 4, and 7 with obvious rotation changes. Therefore, the robustness of this method to image rotation changes needs further verification, echoing the conclusion of Liu et al. (2019).

Given that SIFT is not affine invariant, it cannot achieve good matching performance on aerial and ground images with large viewpoint changes. ASIFT makes the algorithm affine invariant by simulating the affine space. However, the principle of ASIFT assumes that SIFT can completely overcome image scale changes; therefore, the affine space is constructed at the same scale. However, studies (Zhou et al. 2017) have shown...
that in practical applications, when the image scale changes too much, the matching performance of SIFT decreases as the scale difference becomes larger. Therefore, when large scale differences are found between images, e.g., aerial and ground images, the performance of the ASIFT algorithm also drops considerably or even fails.

In addition to NCM and MP, we compared these methods in terms of time efficiency. The average running time of each method on all image pairs is shown in Figure 11. We can see that MeshMatch has the highest time efficiency among all the compared methods. This is because in the MeshMatch method, when the matching of a pair of synthesized and ground images is finished, the matching results can be inversely calculated back to multiple aerial images corresponding to the synthesized image to obtain the matching results of multiple pairs of aerial and ground images. Therefore, in calculating the running time of MeshMatch, if a synthesized image corresponds to n aerial images, the running time is divided by n.

SuperGlue achieves the second-highest time efficiency. This is because in our experiments, the original images are reduced to a size of 1500 × 1000 pixels before they are entered into the matching program, due to the limitation of hardware. We can speculate that if we directly matched the original images, the running time of SuperGlue would increase significantly.

The time efficiency of the proposed method is lower than those of MeshMatch and SuperGlue, but much higher than those of SIFT and ASIFT. In the proposed method, most of the time is spent in the step of local region-based feature-point matching. In this step, we need to perform SIFT matching on each pair of local regions, which is a time-consuming process. Fortunately, it is worth noting that the matching of each pair of local regions is independent of the others, and it is convenient to design parallel programs to improve the efficiency of the proposed method.

**Conclusions and Future Work**

In this study, we proposed a robust feature-point matching method for aerial and ground images by exploiting the constraints of line segment-based local regions. In the proposed method, the ground-view images synthesized from the aerial image-based 3D mesh model were used as a proxy to overcome the global geometric distortion between aerial and ground images, similar to the 3D rendering-based method MeshMatch. However, in consideration of the substantial local geometric deformations, holes, and blurred textures on the synthesized image, the feature-point matching in the proposed method was not performed between the synthesized and ground images. Instead, we designed a geometric relationship-based line-segment matching method to extract line-segment matches from the synthesized and ground images and further generate local-region correspondences between the original aerial and ground images. Afterward, feature-point matching was performed between the original aerial and ground images with the constraints of the local-region correspondences. The proposed matching framework can not only overcome the matching difficulties caused by the viewpoint and scale variations between aerial and ground images but also avoid the negative influence of the local deformations, holes, and blurred textures on the synthesized image during feature-point matching. In our experiments, the proposed method performed better than all the state-of-the-art methods we tested, and increased the average NCM and average MP of the second-best method by more than five times and 40%, respectively, thereby demonstrating its effectiveness. One limitation of our method is that some of the parameters may need to be tuned if the data set is considerably different from normal aerial and ground images. Therefore, in our future work, we intend to use adaptive parameter tuning to improve the general applicability of the proposed method.

**Acknowledgments**

This research was supported by the National Natural Science Foundation of China (41631174, 41971411, and 41971310), the Sichuan Science and Technology Program (2020YFG0083, 2021YFG0028, and 2020JDTD0003), and the Open Fund of State Key Laboratory of Remote Sensing Science (OFSLRSS202004). We would like to thank the anonymous reviewers for their careful reading and comments.

**References**


ASPRS AERIAL DATA CATALOG
“THE SOURCE FOR FINDING AERIAL COLLECTIONS”

HTTP://DPAC.ASPRS.ORG

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

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

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

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

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

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

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

For More Details or To Get Involved Contact:
DAVID RUIZ • druiz@quantumspatial.com • 510-834-2001 OR DAVID DAY • dday@kasurveys.com • 215-677-3119
LEARN
DO
GIVE
BELONG

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

asprs.org