



Calhoun: The NPS Institutional Archive
DSpace Repository

Faculty and Researchers

Faculty and Researchers' Publications

2021

Assessment of Multispectral Imaging System for UAS Navigation in a GPS-denied Environment

Yakimenko, Oleg A.

Monterey, California: Naval Postgraduate School

<https://hdl.handle.net/10945/69769>

This publication is a work of the U.S. Government as defined in Title 17, United States Code, Section 101. Copyright protection is not available for this work in the United States.

Downloaded from NPS Archive: Calhoun



Calhoun is the Naval Postgraduate School's public access digital repository for research materials and institutional publications created by the NPS community. Calhoun is named for Professor of Mathematics Guy K. Calhoun, NPS's first appointed -- and published -- scholarly author.

Dudley Knox Library / Naval Postgraduate School
411 Dyer Road / 1 University Circle
Monterey, California USA 93943

<http://www.nps.edu/library>



NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

**ASSESSMENT OF MULTISPECTRAL IMAGING SYSTEM FOR
UAS NAVIGATION IN A GPS-DENIED ENVIRONMENT**

by

Oleg Yakimenko

October 2021

Approved for public release. Distribution is unlimited

Prepared for: N8 - Integration of Capabilities & Resources

This research is supported by funding from the Naval Postgraduate School, Naval Research Program (PE 0605853N/2098). NRP Project ID: NPS-21-N029-A

THIS PAGE INTENTIONALLY LEFT BLANK



TABLE OF CONTENTS

I. INTRODUCTION.....	6
II. CHOICE OF SUITABLE DCT APPLICATION AND LITERATURE REVIEW.....	4
III. HARDWARE CONSIDERATIONS.....	7
IV. APPLICABILITY OF MS IMAGERY FOR DCT	10
A. UXO IMAGERY COLLECTION AND PREPROCESSING	10
B. CNN ARCHITECTURE DESIGN.....	11
C. DETECTOR TRAINING.....	13
D. PERFORMANCE ASSESSEMNET.....	17
E. SUMMARY	22
V. APPLICABILITY OF MS IMAGERY FOR NAVIGATION	23
A. FEATURE-EXTRACTION ALGORITHMS.....	23
B. GREENSIGHT IMAGERY	25
C. FEATURE EXTRACTION	27
D. FEATURE MATCHING.....	29
E. SUMMARY	37
VI. CONCLUSION	38
LIST OF REFERENCES.....	42
INITIAL DISTRIBUTION LIST	47



THIS PAGE INTENTIONALLY LEFT BLANK



I. INTRODUCTION

Electro-optical (EO) and infrared (IR) imagery are well-known technologies that have been widely used in many applications. This includes military and law enforcement applications involving providing situational awareness both day and night and in low light conditions.

These days, an integrated EO/IR sensor and its associated signal or image processing, tightly integrated in form and function, is a standard payload for manned and unmanned aerial systems (UAS) of Group 1 and higher. Specifically, Group 3 - Group 5 UAS use numerous sensor technologies that span from ultraviolet to far infrared [1].

Critical features of gimballed EO/IR systems are long-range imaging abilities and image stabilization. EO/IR sensors are used to detect and identify targets, track moving targets, and assess threats from a distance and in challenging environmental conditions. Common applications of EO/IR systems include airborne homeland security, combat, patrol, surveillance, reconnaissance, search and rescue programs.

For most gimbal pointing applications, including geo-referencing, the gimbal control system requires position data, which precludes a solution provided by pure inertial measurement unit (IMU) or Heading Reference System (AHRS) and requires a Global Navigation Satellite System (GNSS) -aided option. In order to operate in Global Positioning System (GPS) -denied environments, which is becoming a growing concern as peer-state rivals continue to advance GPS spoofing and denial techniques, new technologies are being explored / developed. The current most promising technologies are Signals of Opportunity (SoOP), magnetic anomaly navigation (MAGNAV) and Vision Navigation (VisNav) [2].

Particularly, VisNav is considered to be one of the most rudimentary forms of navigation and involves building a database of terrain features or landmarks that can then be tracked by onboard sensors in order to calculate vehicle's position, velocity and altitude to provide Precision, Navigation and Timing (PNT) solution. There is a lot of research has been done in this area (e.g. [3,4] and references therein), but the developed prototypes have not necessarily been advanced to the higher technology readiness levels (TRLs) yet. The premise is that as satellite imagery databases and image processing algorithms continue to



advance, visual navigation could likely become more automated and sophisticated. GPS-free vision-based navigation concepts have been studied at the Naval Postgraduate School as well [5,6].

Multispectral (MS) and hyperspectral (HS) imagery is a relatively new technology. The difference compared to the EO and infrared IR sensors is that MS and HS sensors capture reflected light/energy in the number of bands rather than in a single narrow band.

For example, a MS sensor may include 3-10 wide bands (Fig. 1a), while a HS sensor may include hundreds or thousands of much narrower (10-20 nm) bands (Fig. 1b).

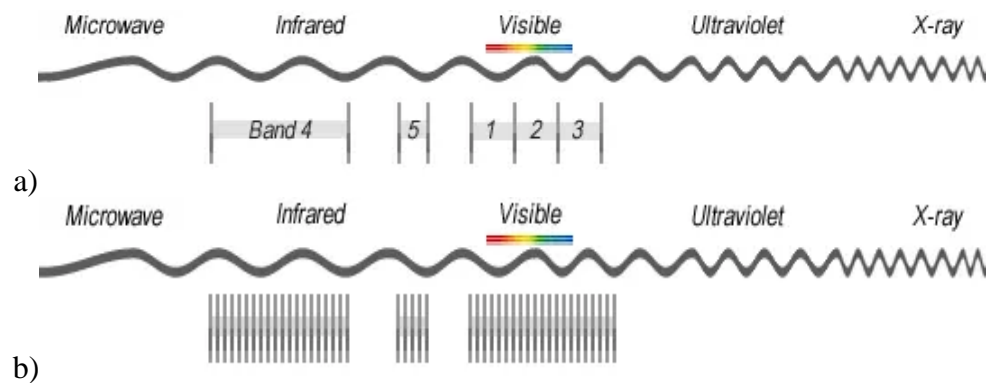


Figure 1. Typical MS (a) and HS (b) sensor bands.

Because MS and HS imaging captures information that cannot be seen with the human eye and presents a more data-rich mosaic for scientists, it has become a highly desired technology for applications within the remote-sensing realm. These applications include crop science, precision agriculture, mining and mineral exploration, petroleum exploration, ecology, disaster mitigation, and others (typically, these applications are most efficiently done from the air).

One of the first HS sensor usages includes NASA’s EO-1 satellite launched in 2000. This satellite carried the hyperspectral sensor “Hyperion” producing 30-meter resolution images in 242 spectral bands (0.4-2.5 um). Other hyperspectral imaging missions from space include PROBA-1 (ESA) in 2001, PRISMA (Italy) in 2019, EnMap (Germany) in 2020, HISUI (Japan) in 2020, and HypSIRI (United States) in 2024 [7].

More compact sensors, compatible with Group 1 UAS include a HS Micro-Hyperspec sensor family compatible with DJI Matrice 600 Pro, FreeFly Alta X, and BFD SE8 [8]. They are available as VNIR, NIR, extended VNIR, and SWIR versions. They

feature up to 369 spectral bands, up to 1,600 spatial bands, and frame rates of up to 450 Hz. HS ranges for these sensors are shown in Fig. 2.



Figure 2. HS ranges for a UAS compatible Micro-Hyperspec sensor family [8].

MS sensors are used in agriculture, environmental monitoring, forestry as well. They enable early detection of issues like disease, water stress, pest infestation, nutrient deficiencies, and more. The specific filters on MS cameras highlight changes in chlorophyll content in plants, which is oftentimes an indicator of disease or stress. Analytics and outputs from MS imagery is used to produce a variety of so-called agricultural indices (Fig. 3) [9]. For example, NDVI (Normalized Difference Vegetation Index) is used to evaluate plant vigor, differences in soil water availability, foliar nutrient content (when water is not limiting), yield potential. CIR Composite (Color Infrared) is used to assess plant health, identify water bodies, assess soil composition and variability in soil moisture. NDRE (Normalized Difference Red Edge) and Chlorophyll Map are used to assess leaf chlorophyll content, plant vigor, stress detection, fertilizer demand, Nitrogen uptake. OSAVI (Optimized Soil-Adjusted Vegetation Index) is used to differentiate soil pixels, account for non-linear interactions of light between soil and vegetation. DSM (Digital surface Model) is used to estimate relative crop volume, identify surface properties, model water flow and accumulation [9].

MS sensors compatible with small UAS (sUAS) include MicaSense RedEdge-MX and Altum. The corresponding wave lengths are shown in Fig. 4. Center wavelengths are as follows (Fig. 5):

- Blue 475 nm
- Green 560 nm
- Red 668 nm
- Red edge 717 nm
- Near infrared 840 nm
- Thermal 11 μm

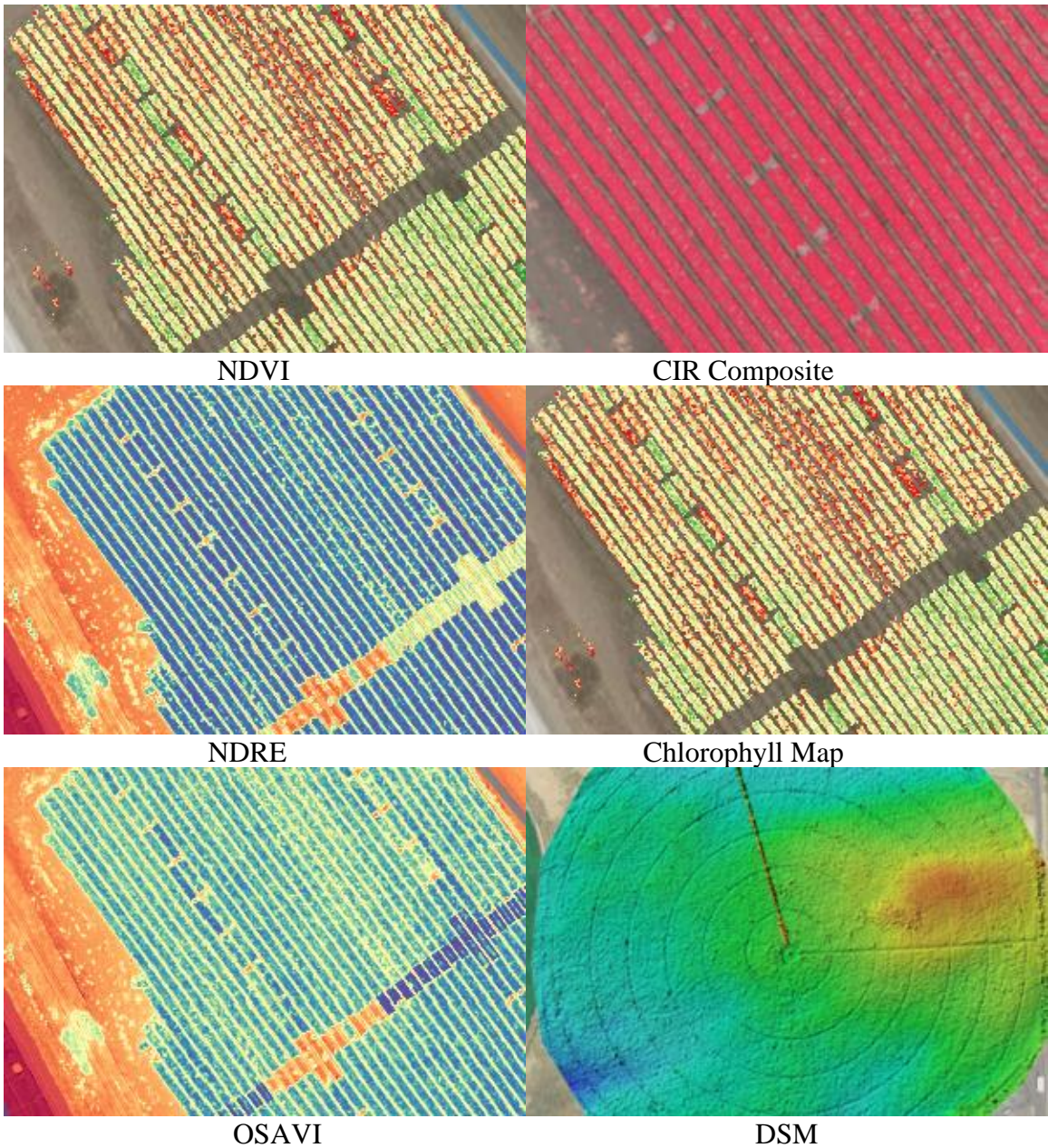


Figure 3. Examples of agricultural indices produced using MS sensors [9].

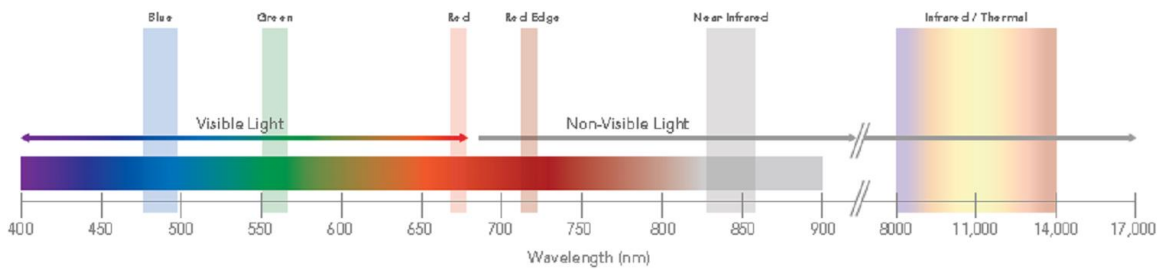


Figure 4. Wavelengths for MicaSense RedEdge-MX and Altum MS sensors [9].

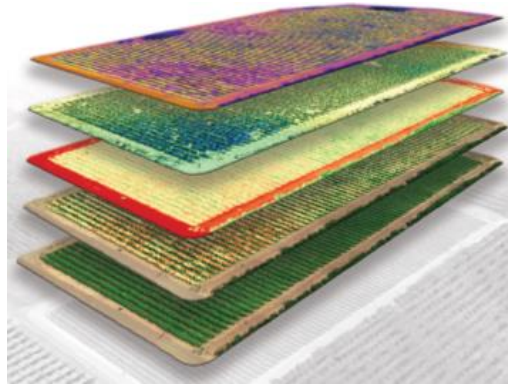


Figure 5. Synchronized thermal and multispectral imagery for MicaSense RedEdge-MX and Altum sensors [9].

The two aforementioned MS sensors were specifically designed for UAS and vegetation mapping and to date have been used by customers in 75+ countries, which includes many researchers. To date, the latter produced over 100 research publications.

This study deals with analysis of feasibility of utilizing the MS and HS technologies in the areas they were not necessarily intended for. Specifically, the objective was to determine whether these technologies can enhance accuracy and precision of object detection (identification), classification and tracking (DCT) that may contribute to a variety of downstream applications including threat detection, forensics, battle damage-assessment, additional/alternative aid to navigation (ATON) in the GPS-degraded or GPS-denied environments. The research questions were formulated as

- Whether using multiple spectral bands has any benefits compared to a standard EO sensor or EO sensor combined with IR sensor? That includes benefits of having a spectral profile of surrounding background area and objects from the standpoint of more reliable/precise DCT.
- What are the limitations of using MS sensors and computer vision / artificial intelligence (CV/AI) algorithms to process data from the standpoint of operating environment, terrain, altitudes, object size and material, time of the day, weather, number of spectral bands, resolution, narrow field of view, addition of a downwelling light sensor)?
- What computational resources would be required to enable DCT capability on board of Commercial Off-The-Shelf (COTS) sUAS?

Two additional questions were

- Whether an onboard MS sensor and available feature-finding and matching techniques can contribute to enabling GPS-free navigation for aerial vehicles?
- What accuracy could be expected when utilizing the non-standard navigational fixes provided by MS sensors?

To address these research questions, this study is organized as follows. Section II discusses two possible suitable DCT applications to be considered. It also discusses the state of the art in these applications. Section III discusses the hardware issues as applied to the current study. Section VI presents the results of collecting and processing imagery for the MS-based detection application (to address the first three research questions). Section V considers the two remaining research questions based on the available data set. Section VI summarizes the results of this study.



II. CHOICE OF SUITABLE DCT APPLICATION AND LITERATURE REVIEW

Originally, this study intended to consider two DCT applications as a suitable problem for assessing applicability and benefits of using MS sensor augmented with CV/AI algorithms. This first one had to do with DCT of unexploded ordinances (UXOs) and the second one - with classifying littoral areas.

UXOs are well-known as a significant threat to warfighters who are serving in the places that used to be battlefields such as the Demilitarized Zone (DMZ) between South Korea and North Korea [10]. The current system to detect UXO based on the ground-based electromagnetic system push equipped soldiers to have to walk the dangerous area in person to detect UXOs. Additionally, the equipment always lacks in the field and the operation wastes a lot of time and effort. In reality, the small-size soldier units still need to cross the dangerous area without any support in many places while they are doing military operations. In the case of DMZ operations or Peace Keeping Operations, soldiers do not need to dig all UXOs in the operation area, but they only need to secure a temporary route to cross. In this case, UXOs on the surface are the most dangerous threat to soldiers.

Currently, most company level infantry troops are already equipped with an sUAS with a vision sensor (camera). The warfighters make use of sUAS to secure their safe route; however, they still need to see and check in person the results broadcasted from sUAS. The object detection technology using a trained artificial convolutional neural network (CNN) – machine learning (ML) and deep learning (DL) – created a new bust in unmanned system industry [11]. This situation motivates an idea to detect UXOs on the surface in real time using EO and MS sensors-based CV enhanced with the ML and DL technology.

Previously, a similar approach was used to develop a sUAS-based system to perform foreign object debris management (FOD) at an airport [12]. The field-testing results showed that FOD management using Group 1 UAS is feasible. The CV algorithms were used to process EO sensor data, but it could have probably been enhanced with MS sensor.

In [13], the feasibility of using DLCNN in inspection of concrete decks and buildings using sUAS was investigated. The authors used images of lab-made bridge decks



taken with a point-and-shoot high-resolution camera to train their model. It was shown that it is feasible to apply DCNN for autonomous civil structural inspections with comparable results to human inspectors when using COTS sUAS and training datasets collected with a point-and-shoot handheld cameras. The authors trained the system with the MathWorks' ALEX-net. They used two modes the architecture that were fully trained mode and transfer learning mode. Comparison of two different modes suggested the appropriate approach to train the UXO detector model.

The rapid mine detection by a drone with a thermal camera was considered in [14]. The plastic mine, PFM-1, which is called the "butterfly mine" is not detected by electromagnetic detectors which are normally used for UXO detection. The authors asserted that available low-cost COTS sUAS equipped with thermal sensor allow accurate assessment of minefield presence, orientation, and potential minefield overlap. The results of the study proved a capability of sUAS to detect PFM-1 in various environments such as temperature variations, moisture content, and burial depth.

A CV algorithm to estimate a vision-based relative position for sUAS was studied in [15]. This study objective was to demonstrate a defensive capability against an unauthorized drone by employing autonomous vision-based pursuit and intercept (via estimating the relative pose of hovering and moving airborne sUAS targets. In conclusion, the research assessed the CV-based range and angular estimation against GPS data. The research also suggested the conceptual design and the choice of a hardware implementation for a COTS-based counter-UAS.

In [16], the detection of underground and near-surface UXOs was attempted using the drone-based transient electromagnetic (TEM) system. The results of a survey proved that the sUAS-based system is safer and more efficient than a ground-based TEM system in UXO detection. However, the TEM-based system was not very effective because of the range and altitude constraints.

In terms of classifying littoral areas, an important topic for planning littoral operations, [17] the effectiveness of a model based upon the Smooth Support Vector Machine (SSVM) algorithm to classify five different coastal classes, object, sand, sky, vegetation and water, was demonstrated in [17]. After training, the model was able to correctly identify 93% of the pixels of the images sent to it for testing. More recently in



2018, the use of remote sensing imagery along with a DCNN to classify eight different coastal landscapes with an accuracy of 95% was demonstrated [18].

Previous efforts to classify bottom types through remote sensing have used a variety of sensors to study this topic. In [19] the combination of LiDAR and Worldview-2 satellite data (using all 8 of the satellite's multispectral bands) were demonstrated to classify three general bottom type classes (Hard, Soft, and SoftHiVeg). A random forest algorithm was used to classify the data. The combination of LiDAR and WV-2 data, with the random forest algorithm used for classification, achieved an accuracy of 76, versus just 59 for LiDAR alone and just 54 for WV-2 imagery alone, i.e., benefiting from two sensor data fusion [19].

In [20], a digital single lens reflex (DSLR) camera was used to capture 370 color (RGB) images and NIR images (with a NIR filter) of natural landscapes. These images were manually segmented and annotated at the pixel level to determine whether they belonged to one of ten classes (Building, Cloud, Grass, Road, Rock, Sky, Snow, Soil, Tree, and Water). A Conditional Random Field was used to train a computer to classify the images, which were separated into two datasets, one dataset had RGB images only and the other dataset had RGB plus NIR imagery. The RGB plus NIR datasets improved the classification accuracy of 7 out of the 10 classes.



Because of hardware limitations addressed in the next section, only the UXO detection problem was ultimately considered.



III. HARDWARE CONSIDERATIONS

For this research, two 5-band MS sensors shown in Table 1 were considered. Due to their price and weight, the RedEdge-MX sensor was chosen.

Table 1. MicaSense Sensors [21].

	RedEdge-MX	Altum
		
Weight	231.9 g (8.18 oz.) RedEdge-MX + DLS 2	406.5 g (14.34 oz.) Altum + DLS2
Dimensions	8.7cm x 5.9cm x 4.54cm (3.4in x 2.3in x 1.8in)	8.2 cm x 6.7 cm x 6.75 cm (3.2 in x 2.6 in x 2.7 in)
External Power	4.2 V - 15.8 V	4.9 V - 25.2 V
Power Input	4.0/8.0W (nominal, peak)	5.5/7.0/10W (standby, average, peak)
Spectral Bands	Blue, green, red, red edge, near infrared (NIR)	EO: Blue, green, red, red edge, near-infrared (NIR) LWIR: thermal infrared 8-14um
RGB Output	3.6 MP (global shutter, aligned with all bands)	High-resolution, global shutter, aligned with all bands
Thermal		FLIR LWIR thermal infrared 8-14um radiometrically
Sensor Resolution	1280 x 960 (1.2 MP per EO band)	2064 x 1544 (3.2 MP per EO band) 160 x 120 thermal infrared
Ground Sample Distance	8cm per pixel (per band) at 120m (~400 ft) AGL	5.28 cm per pixel (per EO band) at 120 m (~400 ft) AGL 81cm per pixel (thermal) at 120m
Capture Rate*	1 capture per second (all bands), 12-bit RAW	1 capture per second (all bands), 12-bit RAW
Interfaces	Serial, 10/100/1000 ethernet, removable Wi-Fi, external trigger, GPS, SDHC	Aircraft: Trigger input, top of frame out, 1 PPS out. 3.3V isolated IO 2x USB 3.0 SuperSpeed ports for WiFi or Ethernet and USB 3.0 Storage.
Field of View	47.2° HFOV (multispectral)	50.2° x 38.4° (multispectral) 57° x 44° (thermal)
Storage	SD Card	USB 3.0 compatible storage devices

* Capture rates vary based on write speed of USB storage device

This sensor was specifically designed to be used with the DJI drones. When RedEdge-MX sensor was procured, it was integrated with the available DJI Inspire drone (Fig.6). The integrated system made its maiden flight, but unfortunately the limitations

imposed by Section 848 of the National Defense Authorization Act for Fiscal Year 2020 (NDAA '20), prohibiting the further usage of DJI drones, hit this study.



Figure 6. DJI Inspire integrated with a RedEdge-MX sensor.

Other UAS were looked at after the aforementioned UAS ban. Specifically, the German Quantum-Systems Trinity F90+ was considered [22]. This high-quality UAS features up to 7 km command and control range, 60...90 min of flight time, wide range of high precision sensors and Live Air Traffic (ADS-B). Even though it turned to be a cost prohibitive solution, one Trinity F90+ owned by a local company was available (Fig. 7a). At that time, this UAS was not compatible with MicaSense sensors, so some design was conducted to allow integration (Fig. 7b).

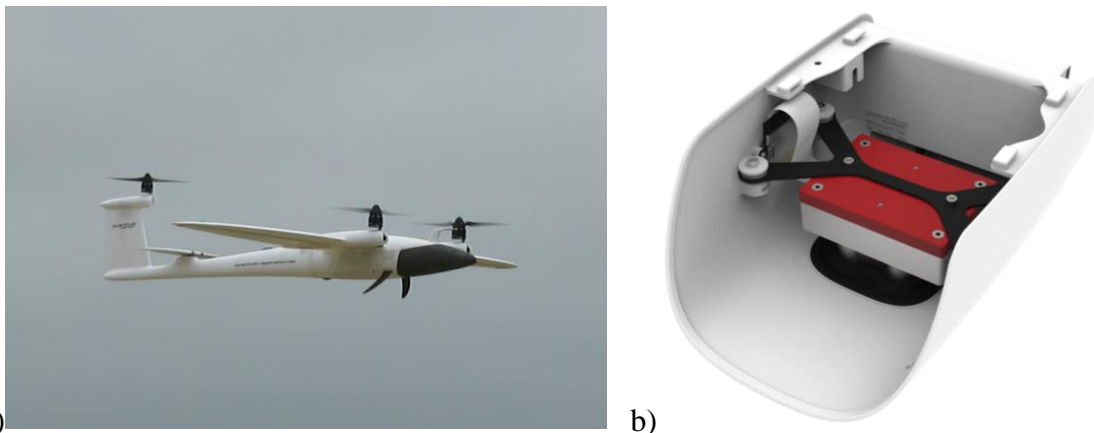


Figure 7. Quantum-Systems Trinity F90+ UAS (a) and RedEdge-MX sensor in the nose bay of UAS (b).

Later on, per Sponsor's suggestion, another US-made UAS was chosen for integration. GreenSight Dreamer UAS [23] can stay in the air for over 60 minutes while carrying GreenSight's MS sensor (Fig. 8). Three such systems were ordered but as of October of 2021, only one was delivered (it was ordered with no camera). Two additional UAS were ordered with a MS sensor, but it turned out that GreenSight only offers a custom 3-band MS camera package (visual, NIR, and thermal). The vendor provided a few samples of imagery from their sensor and this research did some analysis using just these few samples.



Figure 8. GreenSight Dreamer UAS.

Going forward, it should be noted that a Navy COTS UAS Cyber Board waiver (per NDAA '20) was obtained (and is still valid) for both Quantum-Systems Trinity F90+ and GreenSight Dreamer UAS. Moreover, two Certificates of Waiver or Authorization (COAs) were obtained from the Federal Aviation Administration to fly these UAS in Class G airspace between two Restrictive Areas, R-2504 and R-2513, and to the South-West offshore off CA-66 turf runway.

IV. APPLICABILITY OF MS IMAGERY FOR DCT

As mentioned in Section II, the UXO detection problem was chosen to be considered from the standpoint of feasibility / effectiveness of utilizing MS sensors. This section discusses the UXO data collection process, design of CNN, MS detector training, and validation testing conducted on collected data.

A. UXO IMAGERY COLLECTION AND PREPROCESSING

For training the UXO detector, this study acquired a large volume of UXO imagery. Because UAS flights of a MS-equipped platform (Fig. 6) were banned by NDAA '20, this imagery was taken manually, by holding an EO or MS sensor in hands and walking around the planted UXO samples (simulating a low-altitude sUAS flight).

The imagery collection took place at Camp Roberts, CA. Collected data consisted of still images and video clips featuring different UXOs. These included mortar projectiles, hand-grenades, bullets, etc., that had different shape, color, and size. To obtain diverse data, UXO were randomly placed on the ground. Some were placed under the bushes to mimic an operational environment. Figure 9a shows nine different types of UXOs, and Fig. 9b demonstrates a couple of examples of the UXO imagery taken.



Figure 9. UXOs used in this study (a), and examples of imagery taken to mimic an operational environment (b).

In total, 1,225 images and 59 video clips, each about 10 to 20 seconds in length, were taken with an EO sensor (Sony Alpha 6000); and 4,075 images (815 images for each of the five spectrum) were collected using an MS sensor.

The taken images were then preprocessed using the following four-step procedure (for both EO and each band of MS images):

- All images were resized to 416 pix \times 416 pix format - this dimension happens to be the best image size for the YOLOv2 CNN model [24, 25]
- The resized images were manually labeled to represent ground truth data
- Labeled image sets were randomly subdivided into three groups—70% for training, 15% for validation, and 15% for testing
- The number of training images was artificially increased by applying different transformations and tweaking the colormap.

B. CNN ARCHITECTURE DESIGN

To implement the UXO detection, the object detection technique of creating the bounding boxes to mark the predicted object in the input image was applied [26]. After several trials with the ML-based detectors such as the histogram of oriented gradients (HOG), this study chose to apply DLCNN because of its best recognition accuracy for UXO imagery. Among the popular three CNN models for object detection—Faster R-CNN [27], SSD [28], and YOLOv2, the latter one was chosen due to the fastest learning speed [25]. Rather than building new network layers, the pre-trained network structure for the feature extractor was used. Specifically, DarknetReference, Darknet19, ResNet50, and Desnet201 were considered. Finally, due to its advantages in detecting small objects, the ResNet50 was chosen for the UXO detector network.

ResNet50 refers to the Residual Network with 50 layers based on residual learning. In general, as the network goes deeper it becomes more difficult to train the learning feature. Instead of learning features, the ResNet50 is learning the residual inputs from those layers, making it easier to train the network [29]. Through several trials, this study chose the Activation 40 Recertified Linear Unit (ReLU) layer for feature extraction (due to detection performance per training time). The feature extraction transforms the raw data of UXO into numerical information that can be processed. With this approach, the layers



succeeding the feature layers in ResNet-50 were removed. The detection subnetwork along with the YOLOv2 transform and YOLOv2 output layers were added to the feature layer of the base network instead. By doing so, the extracted UXO feature transferred learning by YOLOv2. Table 2 tabulates the developed layer structure for the UXO detector’s network and Figure 10 visualizes it.

Table 2. CNN architecture.

	Component	Deciding Factor
Target Task	Object Detection	Provides the location of the object
Backbone Network	ResNet-50	Small Size of UXO within the Frame
Feature Layer	Activation 40 ReLU	Performance and Training Speed
Detection Network	YOLOv2	Fastest Learning Speed

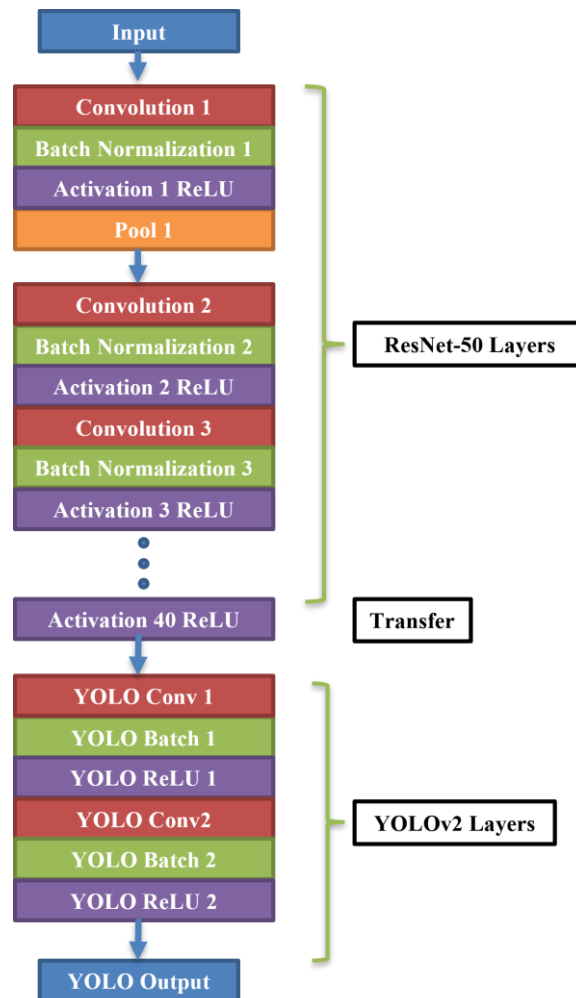


Figure 10. Proposed network layer design for the UXO detector.

C. DETECTOR TRAINING

First, an EO detector was trained (to further compare its performance against that of the MS sensor). The EO detector was trained with the 857 EO images of UXOs (that is 70% of the collected EO UXO images). After several trials with different training options, the number of the anchor boxes (anchor boxes is one of the most important tuning parameters for improving detector's performance) was set as nine, maximum epochs as 20, and mini-batch size as eight. Using a generic laptop (Table 3), the training of the EO UXO detector took about two hours with no significant training loss.

Table 3. Specification of a laptop used in this study.

Microprocessor	Intel® Core™ i7-1065G7 (1.3 GHz base frequency, up to 3.9 GHz with Intel® Turbo Boost Technology, 8 MB cache, 4 cores)
Memory	16 GB DDR4-2666 SDRAM (2×8 GB)
Hard drive	1 TB 5400 rpm SATA
Graphic card	NVIDIA GeForce MX250

Having the EO UXO detector trained, the UXO detection was conducted on the testing and validation imagery sets. To this end, Fig. 11 shows four examples of UXO detection for EO images. As seen, in all four cases shown in this figure, the detection was successful. Figure 12 though shows an example of a failed UXO detection.



Figure 11. Examples of successful detection of single and multiple UXOs.



Figure 12. Example of unsuccessful UXO detection.

What happens with the MS images? To detect UXOs using MS sensor data, this study employed the ensemble method [30], meaning that object detection in each spectral band is conducted by an individual detector separately and then all results are integrated. This is because it is presumed that for a MS sensor, “each spectral image has different characteristics and hence separately detecting objects in each spectral image will help to exploit the feature separately in each spectrum” [30].

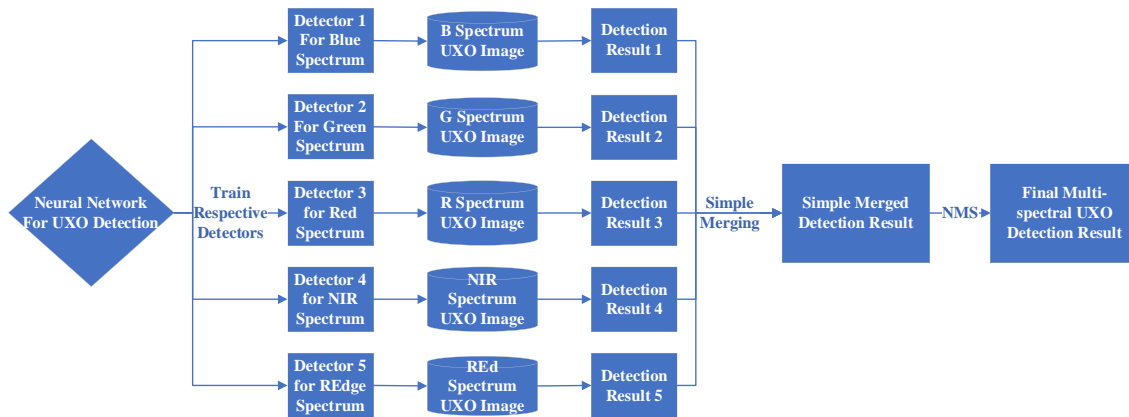


Figure 13. UXO detection workflow for MS imagery.

Following this approach, while only one detector needed to be trained for the EO data set, five detectors were trained for five MS data sets (Detector 1 for the Blue spectrum of the UXO image, Detector 2 for the Green spectrum, and so on) (Fig. 13).

These five detectors were trained in a manner similar to that for the EO detector. Nine anchor boxes, five maximum epochs, and eight mini-batch sizes were used. (The reason for the reduced number of maximum epochs was to reduce the training time without losing overall performance.) Each set of 570 UXO single-spectrum images trained the corresponding detector, which took 20 minutes per spectrum.

Once detectors were trained, the ensemble method was implemented so that each spectrum went into the corresponding detector, that is, the Red spectrum image went into Detector 3, the Blue spectrum image went into Detector 1, and so on (Fig. 11). As a result, each detector detects UXO in its spectrum area and has different detection results from those of other spectrum detectors. Then, the five individual detection results are merged in a two-step procedure: 1) simple merging, and 2) non-maximal suppression (NMS).

The first step is the simple merging that puts the detection results from the five detectors into a single space. The simple merging step corresponding to the case when all detectors detect UXO in each spectrum is shown in Fig. 14.

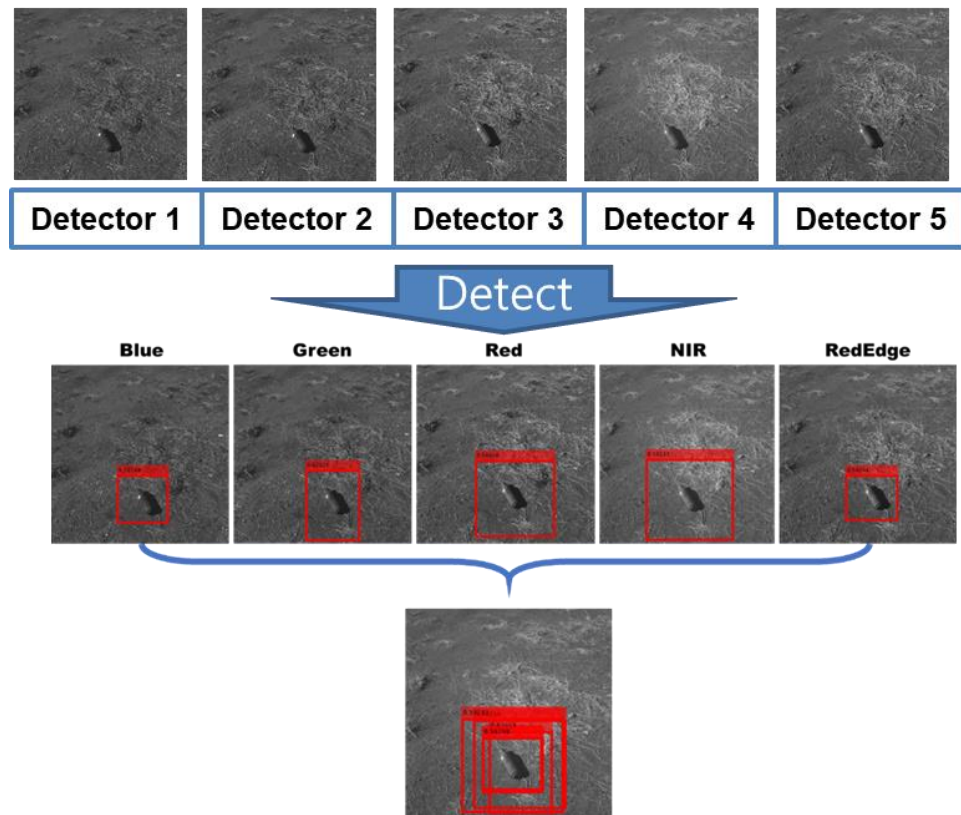


Figure 14. Illustration of a simple merging step.

The next step is suppressing weak detection results and leaving the best detection result by using the Non-maximum Suppression (NMS) method. Following the simple merging step, NMS chooses the best candidate from the detection box with the highest confidence score (to be discussed next) while suppressing all the other overlapped

detection boxes above a certain threshold of intersection over union (IoU) [31]. (IoU indicates the area of overlap over area of the union of the two detection results.)

To prevent the actual overlapped detection result from being recognized as not overlapped, the IoU threshold in this study was set as 0.1. Hence, the redundant detection results with IoU higher than 0.1 can effectively be suppressed. The reason for setting the low IoU threshold relates to the characteristic of a specific MS sensor used in this study - the MicaSense RedEdge-MX sensor has five physically different lenses for five spectra, and the different physical locations of the lenses cause misalignment. As a result, even though each spectral detector successfully detects UXO, the simply merged bounding boxes are misaligned and with a high IoU threshold it makes the system think each spectrum detected a different object. Figure 15 continues an example shown in Fig. 14 and shows a final UXO detection result after the NMS step leaving a single bounding box with the highest score among all detection results.

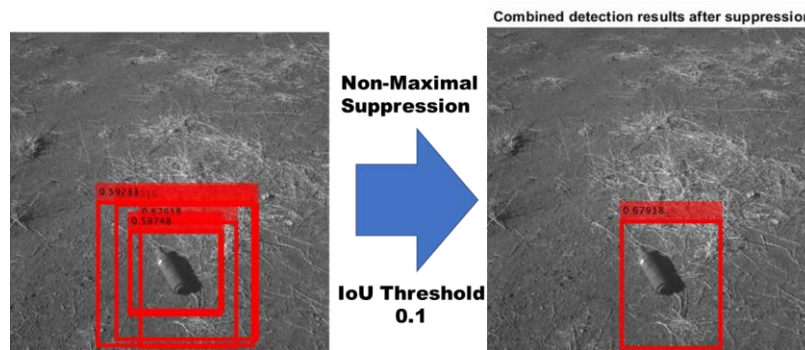


Figure 15. Final detection result with application of NMS method.

The question is then what advantage the ensemble method for a multi-band MS sensor has over a common EO sensor? Data shown in Fig. 16 give a clear answer. In this particular case, while Detectors 1 and 2 do detect UXO, Detectors 3, 4, and 5 fail to do so. The simple merging in this case results in two bounding boxes and the NMS step then merges them into a single bounding box. Hence, each spectral detector complements other detectors which results in a more reliable detection.

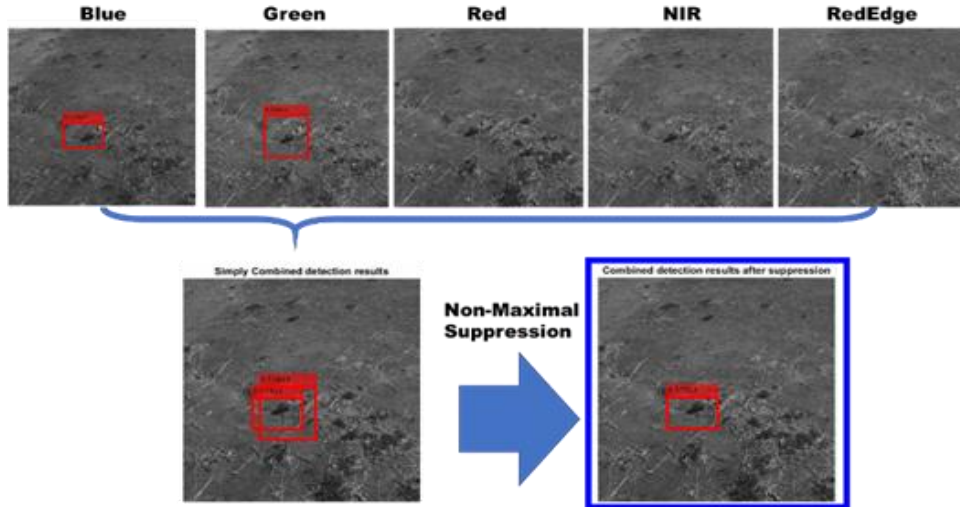


Figure 16. Successful UXO detection with three out of five MS bands failing to do it individually.

D. PERFORMANCE ASSESSEMENT

The performance of a detection algorithm is usually assessed using a confusion matrix. In this matrix, shown in Table 4, each row indicates a test result, and each column indicates an actual state.

Table 4. Confusion matrix for UXO Detector Evaluation [32].

		Actual State		
		True	False	
Test Result	Positive Detection	True Positive	False Positive	Total Positive
	Negative Detection	False Negative	True Negative	Total Negative
		Total True	Total False	

In the UXO detection case, Positive or Negative is determined by whether the UXO detector detects UXO in a given image. True or False is determined by whether the value of the IoU is larger than a certain IoU threshold, indicating how much the detection results and the actual UXO location overlap.

Detection precision measures how accurate the UXO detector’s predictions are based on the ground truth data—in this case, actual UXO location. Hence,

- if the IoU between the prediction and the ground truth is more than or equal to a IoU threshold, the detection is classified as True Positive;
- if IoU is less than a IoU threshold, the detection is classified as False Positive; and

- if a ground truth is present in the picture and the detector failed to detect the object, it is classified as False Negative.

The concept of UXO detector precision, P , is represented by the following relationship [33]:

$$P = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} = \frac{\text{True Positive (Correct Detection)}}{\text{Total Positive (All Detections)}} \quad (1)$$

Another measure, recall, r , measures how well the detector finds all positives based on the ground truth data. The recall of the UXO detector evaluation is mathematically defined as

$$r = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} = \frac{\text{True Positive (Correct Detection)}}{\text{Total True (Actual UXO)}} \quad (2)$$

For assessing performance of UXO detector, the IoU threshold was set as 0.5 for the EO-based detection and 0.4 for the MS-based detection. The reason the MS detector has a lower threshold is that MS are not aligned with each other as discussed earlier. This misalignment can cause a phenomenon in which a True Positive—correct detection result—is recognized as a False Positive—wrong detection result. Setting a lower threshold for the MS detection can prevent the actual True Positive from being evaluated as False Positive.

Once the test data set is run through the trained UXO detector, the detection results are ranked in the descending order based on the predicted confidence level. To visualize the Precision and Recall values of the detection results, the Precision-Recall (PR) curve can be plotted, which indicates the precision of the trained detector at different recalls [34].

Average precision, AP, then computes the average precision value for all recalls, i.e., finds the area under the PR curve. In the PASCAL Visual Object Classes (VOC) Challenge [35], the shape of the PR curve is calculated as the mean precision at a set of 11 equally spaced recall levels $[0, 0.1, \dots, 1]$. The 11-point interpolated AP is calculated as

$$\text{AP} = \int_0^1 P(r) \approx \frac{1}{11} \sum_{r \in \{0, 0.1, \dots, 1\}} P_{\text{interp}}(r), \text{ where } P_{\text{interp}}(r) = \max_{\tilde{r} \geq r} P(\tilde{r}) \quad (3)$$

The calculated AP would thus be between 0 and 1, and it is presumed that the higher the AP, the better the detector. The AP value is commonly used as an ultimate metric to evaluate the effectiveness of trained detector.



To this end, Fig. 17, shows the results of computing the AP value for the EO detector. As shown in the table on the left in Fig.17, there were 193 detection instances in the 184 test images. For these 193 instances, the values of P and r were computed based on the ground truth data using Eqs. (1) and (2). The table was then ranked by the Precision column. The PR curve shown on the right was produced then. Using Eq.(3), the AP value (the area under the PR curve computed using an 11-point interpolant), was estimated to be 0.774.

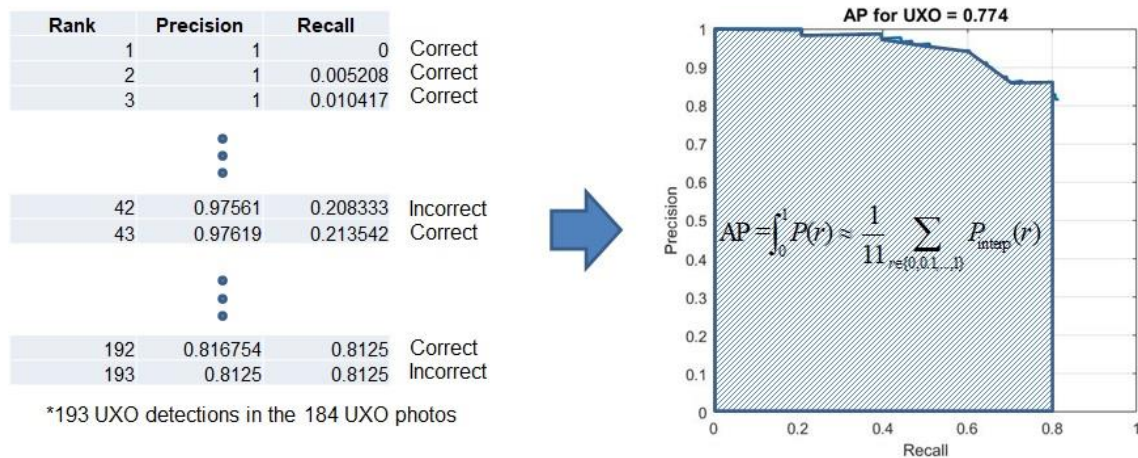


Figure 17. The AP value computation for the EO detector.

For MS detection, the respective detectors were evaluated with the 122 test images for each spectrum. The PR curves for the respective detectors are presented in Fig. 18 along with the AP values spread between 0.484 and 0.592 (shown in the legend). It should be noted that the PR curve for Blue, Green, and Red detectors include the (0, 0) point meaning that these detectors failed to detect UXO in some images (like it was shown in Fig. 16). On the contrary, the PR curve for NIR and Red edge detectors does not pass through the (0, 0) point indicating that these detectors did find some suspected UXO within each image even if there were no UXO (False Positive detection). Figure 19 shows the integrated PR curve for the MS-based UXO detection after applying the two-step merging procedure described earlier (Figs. 13, 15). The AP value for the integrated PR curve happens to be 0.871, i.e., 13% higher than 0.774 for the EO PR curve.

All AP values are shown in Table 5. As seen, while the evaluation results for all single-band detectors were lower than that of the EO sensor, when combined, the AP value for the MS sensor exceeded that of the EO sensor.

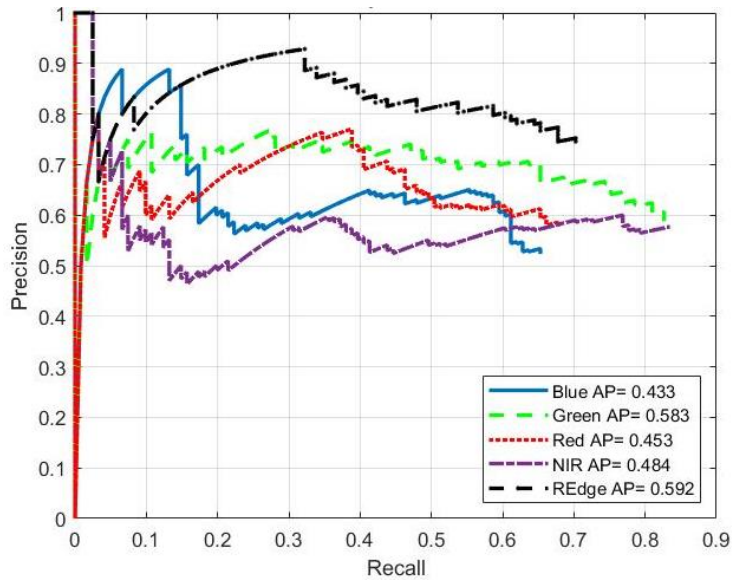


Figure 18. Individual Precision-Recall curves for five MS detectors.

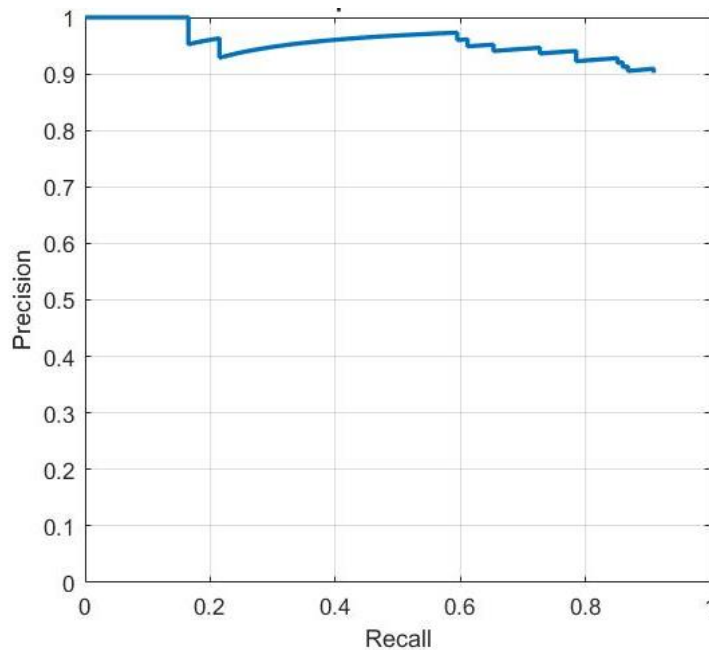


Figure 19. The overall Precision-Recall curve for MS-based UXO detection.

Table 5. Comparison of EO and MS AP values for UXO detection.

Individual Spectrum Detector		Average Precision	
EO Detector		0.774	
Blue Spectrum	Multi-spectral Detector	0.433	0.871
Green Spectrum		0.583	
Red Spectrum		0.453	
Near Infrared Spectrum		0.484	
Red Edge Spectrum		0.592	



Figures 20 and 21 show the results of one more experiment conducted with the EO data. The premise was that the size of the bounding box while creating a set of the ground truth data (which can be controlled during a manual procedure) has an effect on the quality of the detector. Figure 20a shows the results of UXO detection using a detector trained on a tight bounding box set of the ground truth data as opposed to the Fig. 20b showing the case of using a detector trained on a loose bounding box set of the ground truth data. Figure 21 shows the PR curves in both cases. Overall, the effect happens to be negligible (1% improvement for the tight bounding boxing).

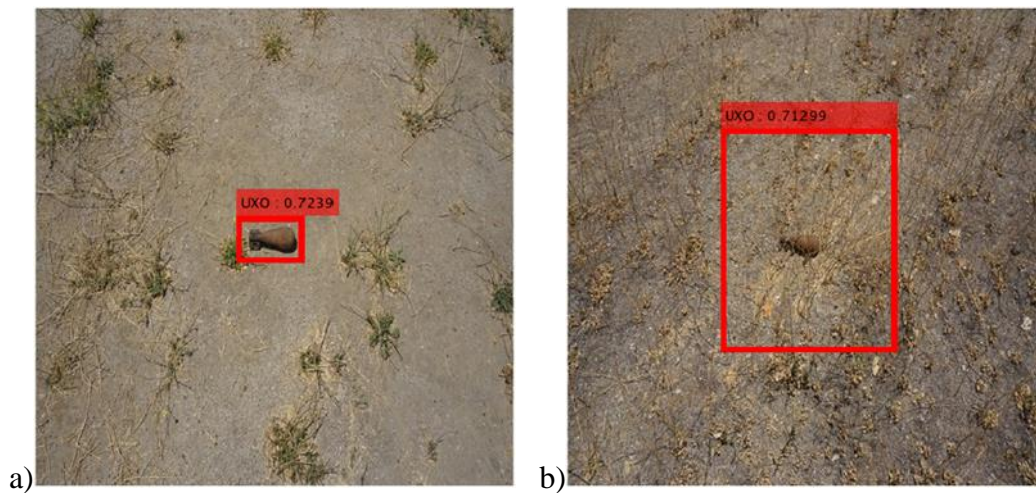


Figure 20. UXO detection using a tight (a) vs loose (b) bounding boxing.

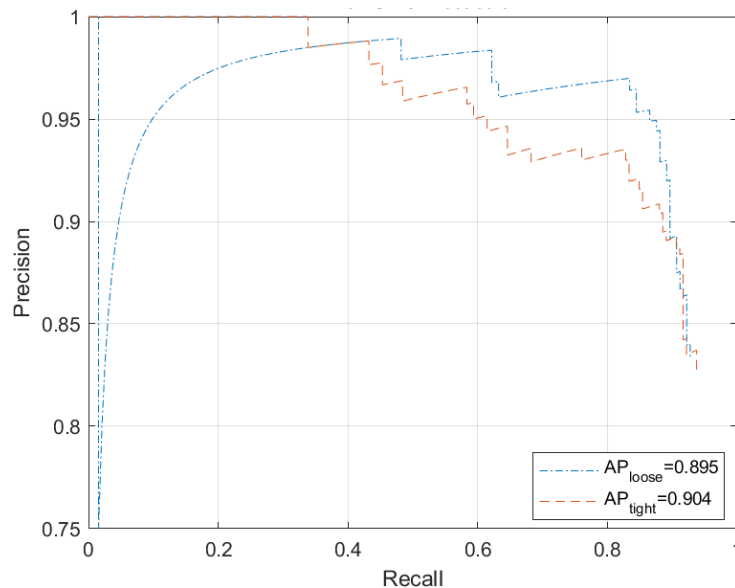


Figure 21. The PR curves for the loose and tight bounding boxing while the EO detector training.

E. SUMMARY

To summarize:

- The AI-based UXO detection system seems to be a good representative application where MS sensor could reveal its potential to contribute to the solution of DCT problems
- As anticipated, the UXO detection capability by individual spectrum detectors is lower ($AP < 0.6$) compared to that of EO sensor ($AP = 0.774$) (Figs. 17, 18). However, it was found that they are complimentary to each other
- By applying the two-step integration process, the overall UXO detection capability of MS detector ($AP = 0.871$) exceeds that of EO sensor by about 13% (Figs. 18, 19, Table 5)
- Using more than one spectral band makes a detection process more reliable. For example, while in some cases Blue and Green detectors were able to detect UXO, other band detectors were not (Fig. 16). In some other cases, only the NIR detector was able to detect UXO while all other detectors failed to do so
- All detectors detecting UXO in their own spectrum band feature a different size of the bounding box and different detection confidence score (i.e., some spectrum detectors detect UXO more precisely than others). For instance, in some cases, the Blue and Green detectors detected UXO more precisely than other detectors (Fig. 14). In some other cases, it was the Red and NIR detectors that detected UXO more precisely than the others
- The limitations imposed by NDAA'20 precluded from exploring a variety of environmental effects (including time of the day, cloudiness, altitude of flying, etc.) as was originally planned



V. APPLICABILITY OF MS IMAGERY FOR NAVIGATION

This section considers applicability and benefits of using the MS sensor to aid GPS-free navigation. It starts from providing some basics on the image feature extraction and matching and proceeds with analysis of a limited set of 2-band MS data provided by GreenSight.

A. FEATURE-EXTRACTION ALGORITHMS

As explained in Ref. [5], the vision-based navigation relies on a feature detection capability. When the camera makes a small motion, it is reasonable to assume that there exists some geometric transformation that maps some features / points found in the first image to the corresponding features / points found on a consecutive image. Different feature-extraction algorithms explored in this study included

- cornerPoints - this is a single-scale corner detection algorithm, applicable for point tracking, image registration with little or no scale change, corner detection in scenes of human origin, such as streets and indoor scenes. To determine the corners, one of the following three different approximate metrics are used:
 - FAST (Features from Accelerated Segment Test) [36]
 - MinEigen (Minimum Eigenvalue algorithm) [37]
 - Harris (Harris–Stephens / Shi and Tomasi algorithm) [38]
- ORB (Oriented FAST and rotated BRIEF) [39] - this is a multiscale corner detection algorithm handling changes in rotation (BRIEF stands for Binary Robust Independent Elementary Features). This method was developed to provide a fast and efficient alternative to a scale-invariant feature transform (SIFT)
- BRISK (Binary Robust Invariant Scalable Keypoints) [40] - this is another multiscale corner detection algorithm. Unlike single-scale corner detection approach, it handles changes in both scale and rotation
- SURF (Speeded-Up Robust Features) [41] - this is a multiscale blob detection algorithm with scale and rotation changes
- KAZE (derived from the Japanese word kaze, 風, meaning breeze or wind) [42] - this is another multiscale blob detection algorithm
- MSER (Maximally Stable Extremal Regions) [43-45] - this is a multiscale regions of uniform intensity detection algorithm for registration, wide baseline stereo calibration, text and object detection. It handles changes in scale and rotation. It is more robust to affine transforms in contrast to other detectors



- HOG (Histogram of Oriented Gradients) [46] – this is a gradient-based feature detection algorithm that uses a grid of uniformly-spaced rose plots, showing the distribution of gradient orientations within a HOG cell.

As applied to all five bands of MicaSense RedEdge-MX sensor, there is more than sufficient number of features found by most of the aforementioned feature-extraction algorithms (of course, each algorithm finds a different number of features). From the efficiency standpoint, Figure 22 shows the computational time required to find a single feature (note, the computational time shown is for an uncompiled code, - for a compiled code it is expected to require two orders of magnitude less time). As seen from this figure, Thermal, Blue and Red spectra require less computational resources than other spectra. The Harris feature detection algorithm appears to be the most efficient, followed by MSER and SURF algorithms.

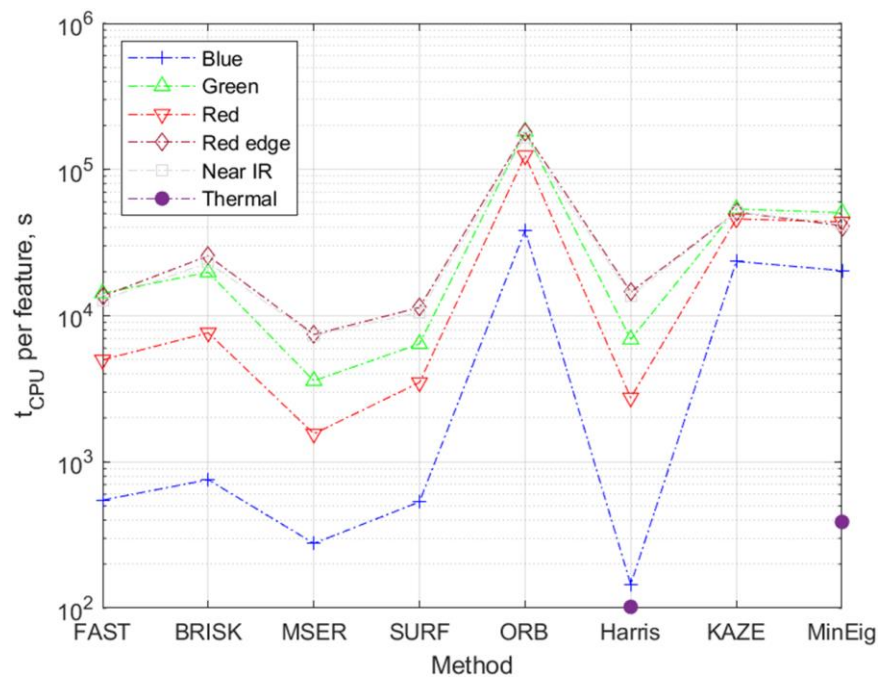


Figure 22. Computational time required to find a single feature point using different feature detection methods.

Once all features are found, a similarity or affine transform can be applied to compute the transition and rotation between the frames. To validate the developed algorithm, Fig. 23 shows a comparison conducted to compute a transformation between the different lenses of the MicaSense RedEdge-MX sensor. For example, the first row of images shows the difference between the Blue and all other bands

excluding Thermal (Green, Red, Red edge, NIR). Unfortunately, the preliminary results conducted on the very limited imagery obtained during the maiden flight of the DJI Inspire drone were not enhanced any further (because of NDAA'20 ban prohibiting further flights).

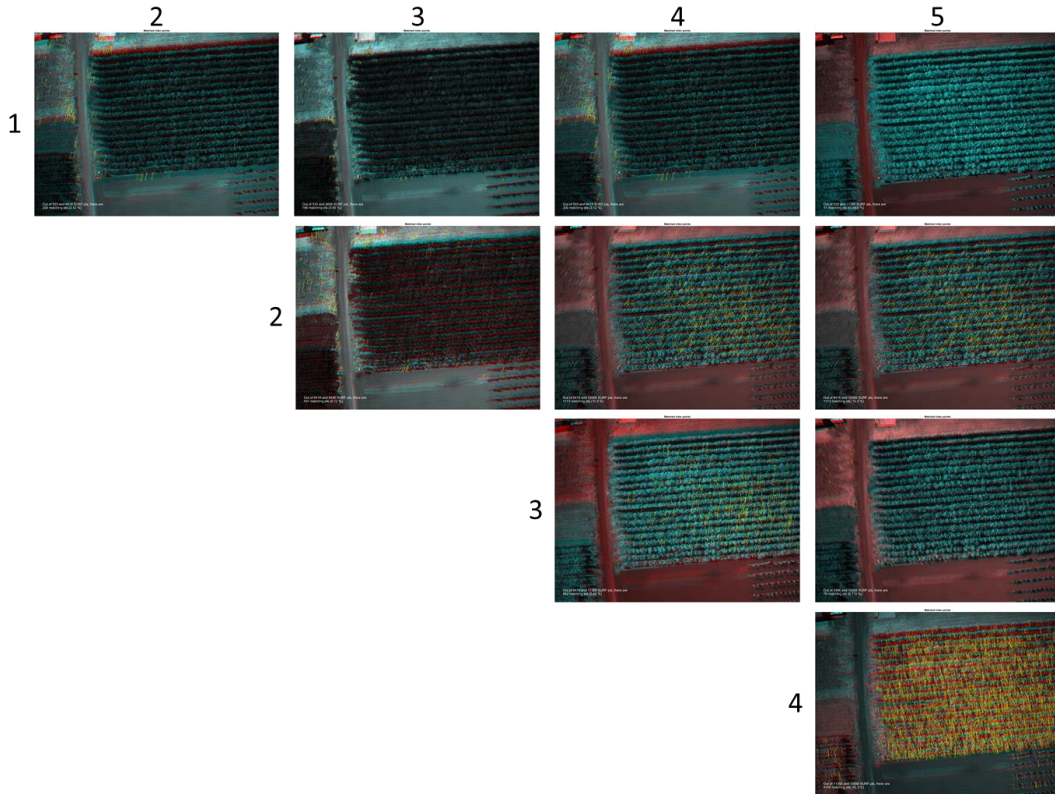


Figure 23. Pairwise comparison (difference) between the different bands of MS sensor.

B. GREENSIGHT IMAGERY

As explained in Section III, GreenSight provided some imagery collected using their own 3-band MS sensor. Specifically, the set of 14 EO and 14 NIR images (taken about 10s apart) were presented for analysis (Table 6). These 2×14 images, taken about 10m above the ground level, are shown in Fig. 24. Figure 25 presents a bird's-eye view of a GreenSight Dreamer UAS trajectory when these images were taken (14 image-taking locations are depicted with the green circles). Even though the fact that provided images were taken 10s apart violates a premise of small motion (Ref. [5] shows that number of the matching points between the two images decreases rapidly with a decrease of the sampling rate), some limited analysis conducted on these two sets of 14 images allowed to confirm some expected trends, as discussed next

Table 6. Data provided by GreenSight.

telem_timestamp	img_timestamp (1)	lat (2)	lon (2)	alt_msl (3)	alt_rel (3)	roll	pitch	yaw	img_idx	img_filename (4)
2021-04-26T15:06:26.115	1619474786003000	42.3005191	-71.0888518	52.95	10.57	-1.67	3.13	165.19	1	IMG_0000 or IMG_0001
2021-04-26T15:06:30.369	1619474790202000	42.3004466	-71.0887361	53.05	10.67	8.22	14.51	170.38	5	IMG_0002
2021-04-26T15:06:31.852	1619474791603000	42.3004356	-71.0887361	53.35	10.97	10.65	17.46	169.13	6	IMG_0003
2021-04-26T15:06:34.703	1619474794603000	42.3004567	-71.0888035	53.05	10.67	13.78	11.54	170.09	7	IMG_0004
2021-04-26T15:06:42.140	1619474801802000	42.3005104	-71.0891228	52.08	9.7	-0.79	6.14	164.63	15	IMG_0005
2021-04-26T15:06:43.547	1619474803403000	42.3004990	-71.0891321	52.41	10.03	-5.39	5.82	166.16	16	IMG_0006
2021-04-26T15:06:54.085	1619474814003000	42.3003715	-71.0887547	52.44	10.06	11.32	7	171.07	26	IMG_0007
2021-04-26T15:06:55.564	1619474815403000	42.3003517	-71.0887597	52.27	9.889999	12.17	14.66	170.73	27	IMG_0008
2021-04-26T15:07:05.073	1619474824803000	42.3004542	-71.0891266	52.13	9.75	0.64	0.14	167.31	36	IMG_0009
2021-04-26T15:07:06.082	1619474826002000	42.3004583	-71.0891449	52.14	9.76	-5.32	-3.37	171.54	37	IMG_0010
2021-04-26T15:07:09.828	1619474829603000	42.3004054	-71.0890968	52.31	9.929999	-3.06	4.57	166.55	38	IMG_0011
2021-04-26T15:07:17.101	1619474837002000	42.3003148	-71.0888067	52.53	10.15	8.99	10.2	163.25	46	IMG_0012
2021-04-26T15:07:26.068	1619474846002000	42.3003402	-71.0890475	52.49	10.11	-3.57	1.36	171.48	51	IMG_0013
2021-04-26T15:07:35.541	1619474855303000	42.3003424	-71.0890550	68.43999	26.06	3.41	5.52	168.71	52	???

- (1) microseconds since unix epoch
- (2) lat & lon recorded in decimal degrees
- (3) altitudes recorded in meters
- (4) Joe's guess based on visual inspection of images and telemetry playback

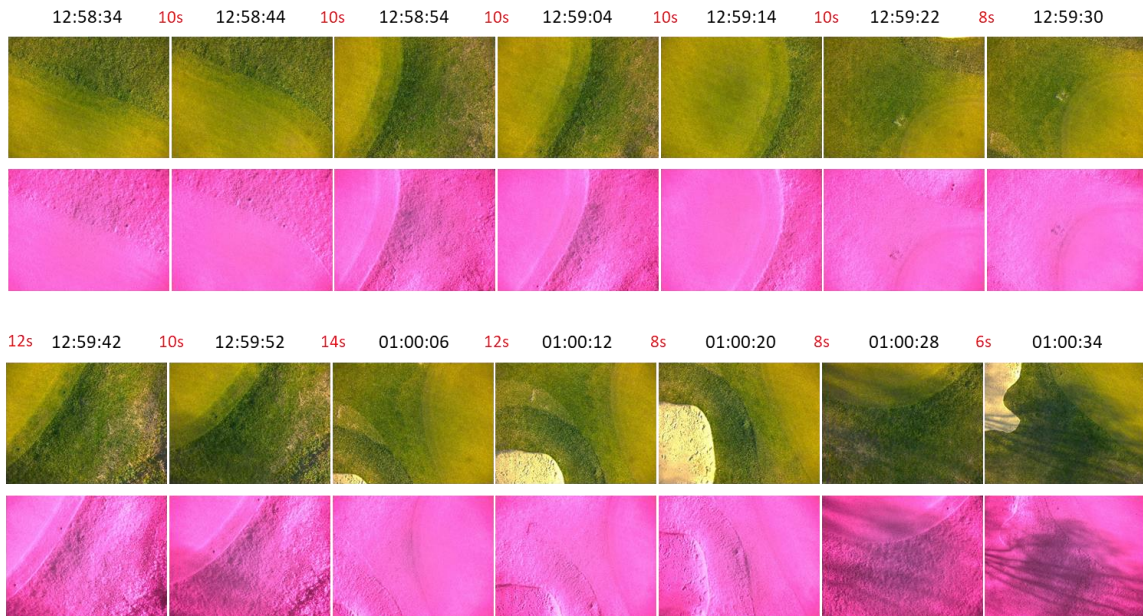


Figure 24. Available GreenSight imagery (4032 pix × 3040 pix).



Figure 25. Bird's-eye view of a GreenSight Dreamer trajectory with locations of still imagery taken.



C. FEATURE EXTRACTION

First, Figs. 26 and 27 visualize the features found on EO and IR images using six different feature extraction methods mentioned earlier in this section. As seen from these figures and also from Fig. 28a, the number of the features found within each image is huge (compared to that of when a similar EO sensor was flown at 2,000'-8000' AGL [5]). The number of the features found within a NIR image (Fig. 27) is on the order of magnitude lower than that of the corresponding EO image (Fig. 26). Figure 28b shows the required computational resource (computation time) to find all the features for each method. Figure 29 shows a required computation time per feature.

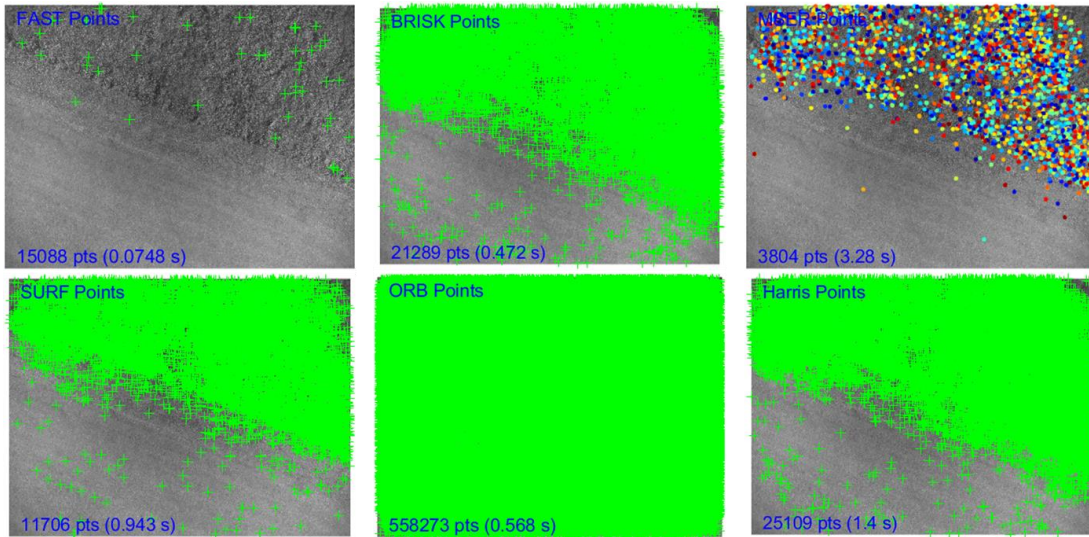


Figure 26. Feature extraction for an image taken with an EO sensor.

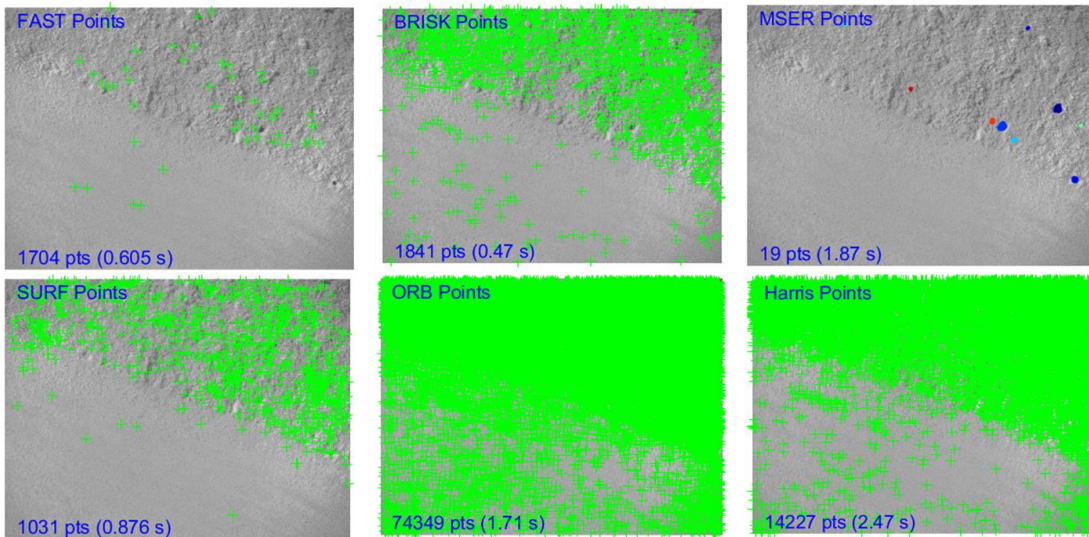


Figure 27. Feature extraction for an image taken with a NIR sensor.

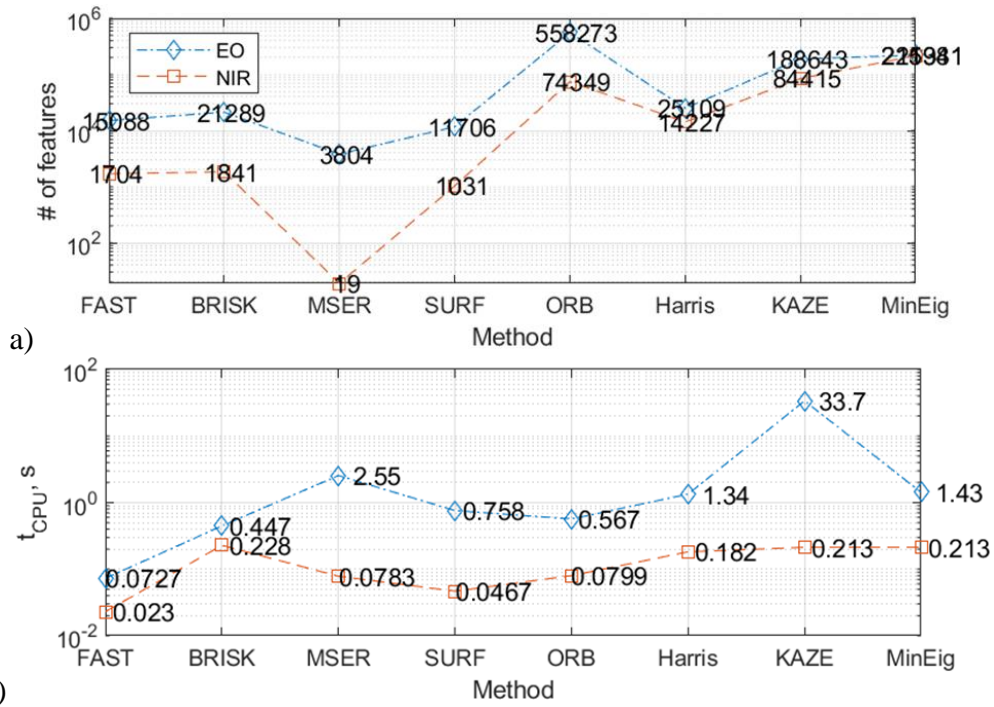


Figure 28. Number of features (a) and computational resources required to find a single feature (b) using different feature detection methods.

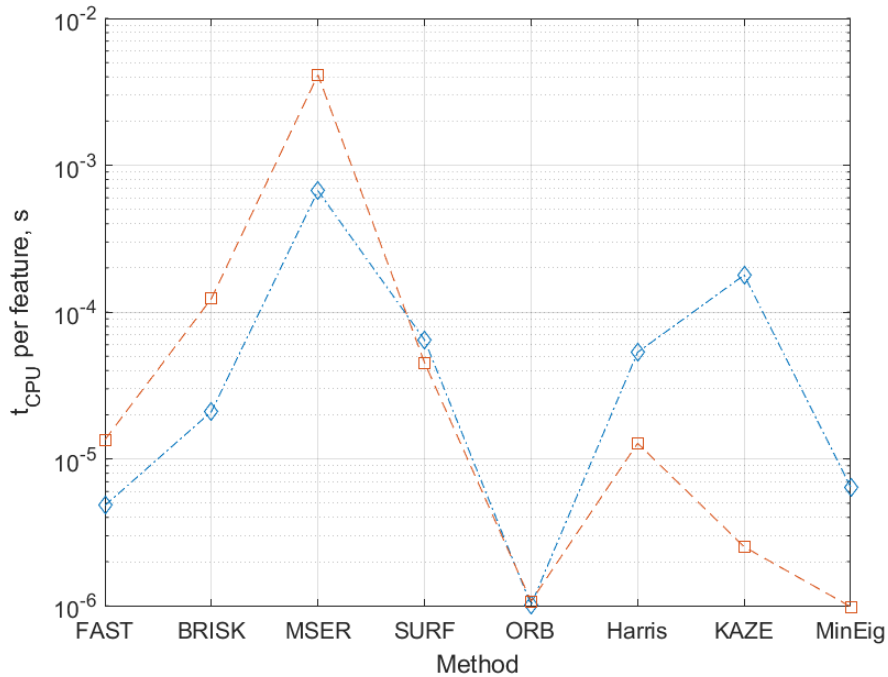


Figure 29. Computation time per feature for different feature-finding methods.

Figure 30 combines data shown in Fig. 28 a, b. From the effectiveness standpoint, the more features the feature-finding method finds and the faster it does it, the better (i.e., one should look at the method that delivers performance appearing the closest to the

bottom-right corner of Fig. 30). As seen from Fig. 30, overall, the FAST and ORB methods are probably the most efficient ones for both EO and NIR images (it is worth mentioning that the Ingenuous helicopter on Mars uses the FAST algorithm as well).

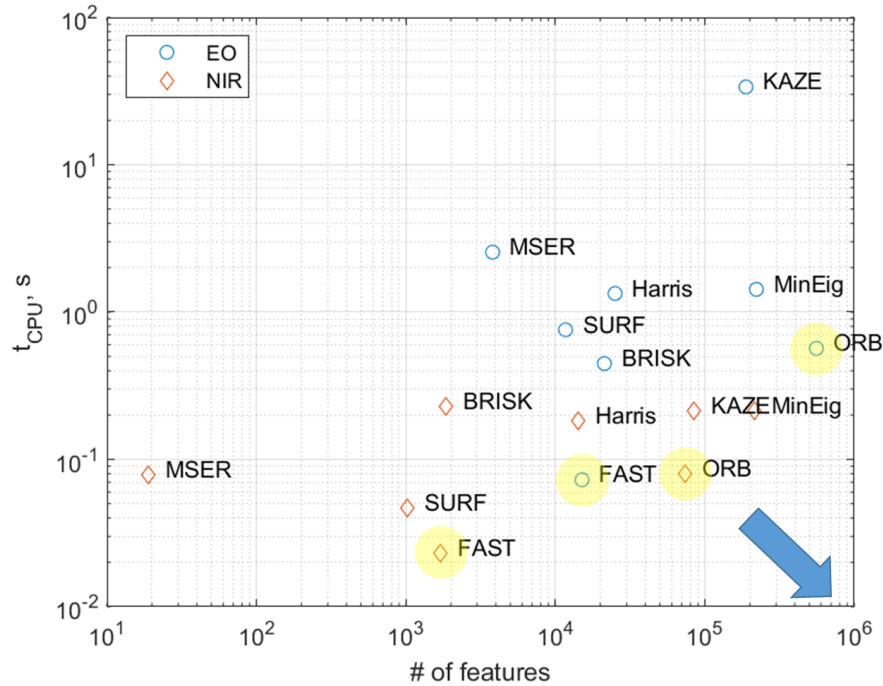


Figure 30. Performance overview of explored feature-finding methods.

D. FEATURE MATCHING

Once the sets of good features are found in two consecutive images, they should be matched, i.e., the subset of those features belonging to both images needs to be found. Out of these found matching features, the so-called inliers (i.e., the matches supposedly belonging to the same transformation) should be further determined. The process of excluding outliers is based on using the M-estimator SAmple Consensus (MSAC) algorithm. The MSAC algorithm is a variant of the Random Sample Consensus (RANSAC) algorithm. (It should be noted that running the same outlier exclusion algorithm multiple times will not lead to the identical result. This is because of the randomized nature of the MSAC algorithm. Having more features in each image to begin with may be a mitigating factor.)

Even though the number of features found in the NIR images is lower than the number of features found in the EO images (which in general may be crucial if the image

capture rate is low and/or UAS flies fast), it was still sufficient to do feature matching to determine transformation between two “consecutive” images (10s apart). Double quotation marks are used here to emphasize that two images 10s apart are not necessarily truly consecutive, because the scene beneath may change drastically and as a result no matching features would be found at all. Looking at Fig. 24, it is clear that while the first two images feature about the same scene (UAS was essentially hovering above the same point with no rotation), transitioning to the third image may result in no matching features at all). Switching from Image 5 to Image 6 represents a complete change of a scene. The same is true for Images 7-7, 9-10, 12-13, 13,14.

To elaborate on this issue further, Fig. 31 shows pairwise feature matching between the first four images in a sequence of images provided by GreenSight. Row wise, Fig. 31 shows three steps – feature matching, the process of excluding outliers using the MSAC algorithm, and applying similarity transformation derived from the matching inliers.

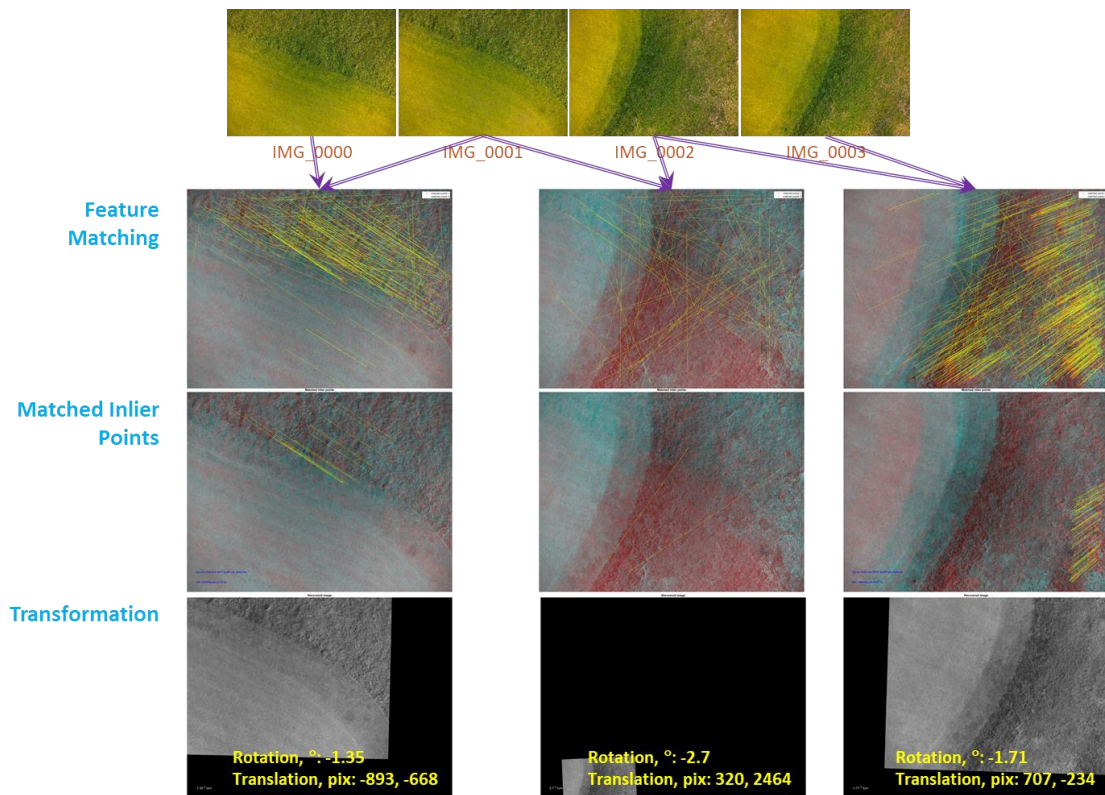


Figure 31. Computing and applying similarity transformation based on matching features obtained using the SURF algorithm (for four first images, 10s apart).

For the first column and third columns of images, showing all steps when comparing the first two and the last two images, showing pretty much the same scene, only 15...20% are found to be the matching features (row 1). Out of these matching features only a small fraction is found to be inliers (row 2). This accounts to only 2...6% of the number of individual features found on each image. Yet, because of the large number of original features, this low percentage yields about 250...700 inliers to compute a transformation (translation and rotation). For example, the transition from the 1st image to the 2nd involves UAS movement down to the right (-893pix, -668pix) and slight counterclockwise rotation (by 1.35°). The transition from the 3rd to the 4th image - movement down to the left (707pix, -234pix) and slight counterclockwise rotation (by 1.71°).

As opposed to the first two and last two images, switching between the 2nd and 3rd images in Fig. 31 causes a problem. As seen from the image located in the 1st row and 2nd column on the image matrix, the found matches happen to be all over the place. Out of these few, only 2 (!) matches are determined to be inliers (2nd row in the 2nd column). Even though a minimum number of matching pairs for the similarity transformation is two, the computed transformation cannot be trusted - the two inliers clearly suggest two contradictory transitions.

Figures 32 and 33 demonstrate a similar analysis performed on three images apparently featuring the same scene (Images 10-12 in Fig. 24). In this case, the feature matching for both EO images (Fig. 32) and NIR images (Fig. 33) yields reasonable results. The number of matching features, even for the images that are 20s apart, is big enough to find a good set of colinear inliers to produce a similarity transform. To this end, Fig. 34a shows the number of matching features found in each pairwise comparison (2nd image vs 3rd image, 1st image vs 2nd image, and 1st image vs 3rd image), and Fig. 34b shows a percentage of matching features (out of the maximum number of individual features in each pair of images). Consequently, Fig. 35a shows the number of inliers found for the same pair, while Fig. 35b shows a percentage of inliers (out of the maximum number of individual features in each pair of images).



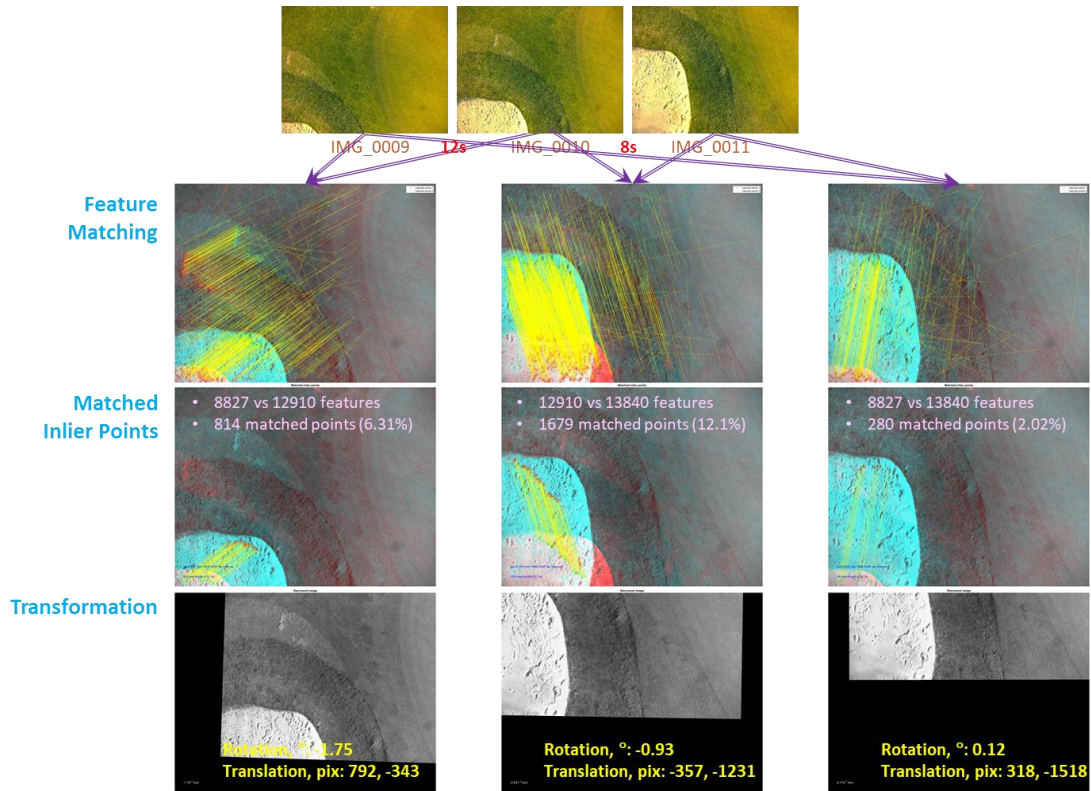


Figure 32. Three “consecutive” EO images (10s apart) featuring about the same scene.

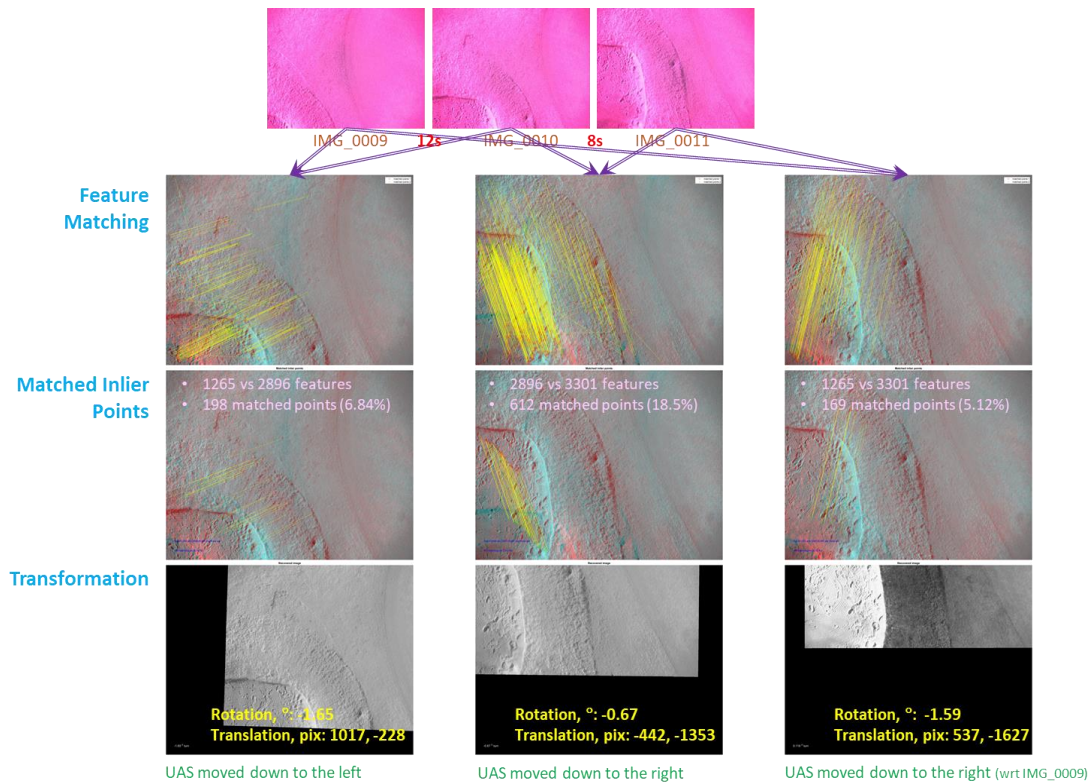


Figure 33. Three “consecutive” IR images (10s apart) featuring about the same scene.

As seen from Figs. 34 and 35, even if UAS hovers above the same point, the time between the two images are taken has a strong effect. It is only about 1% of the total number of features found in each individual image that contributes to computing the transformation. That is why it is important to keep the number features found in each individual image substantially big (in other words, it is important to have a feature-rich terrain and good sensor resolution). Another observation is that while the number of inliers for the NIR images is about three times lower than that of the EO images (Fig. 35a), percentagewise (Fig. 35b) the NIR images exhibit a larger fraction of inliers.

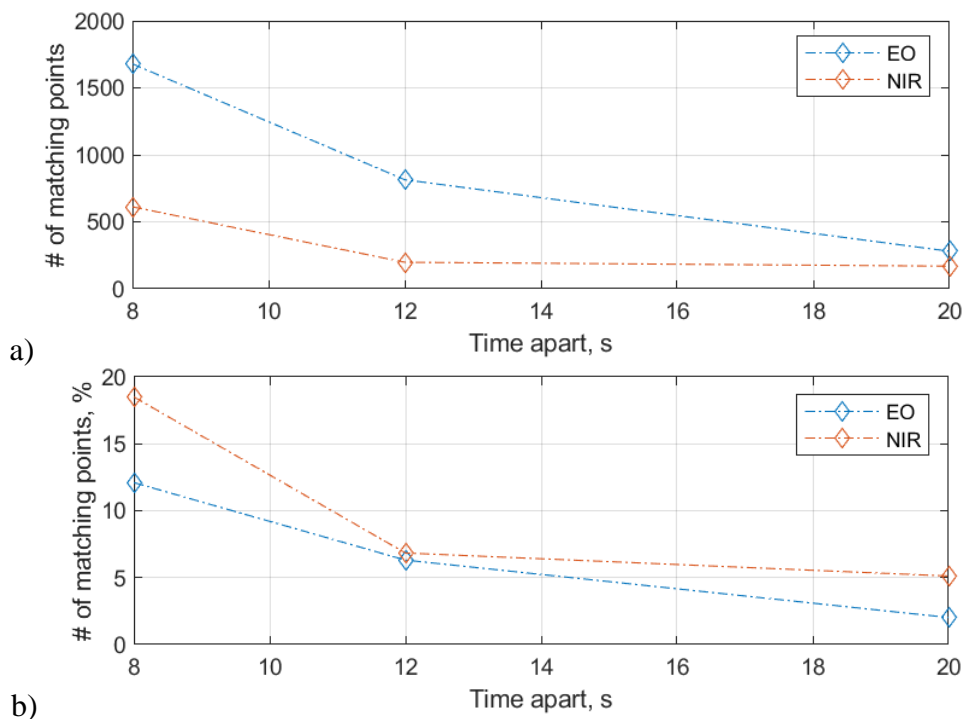
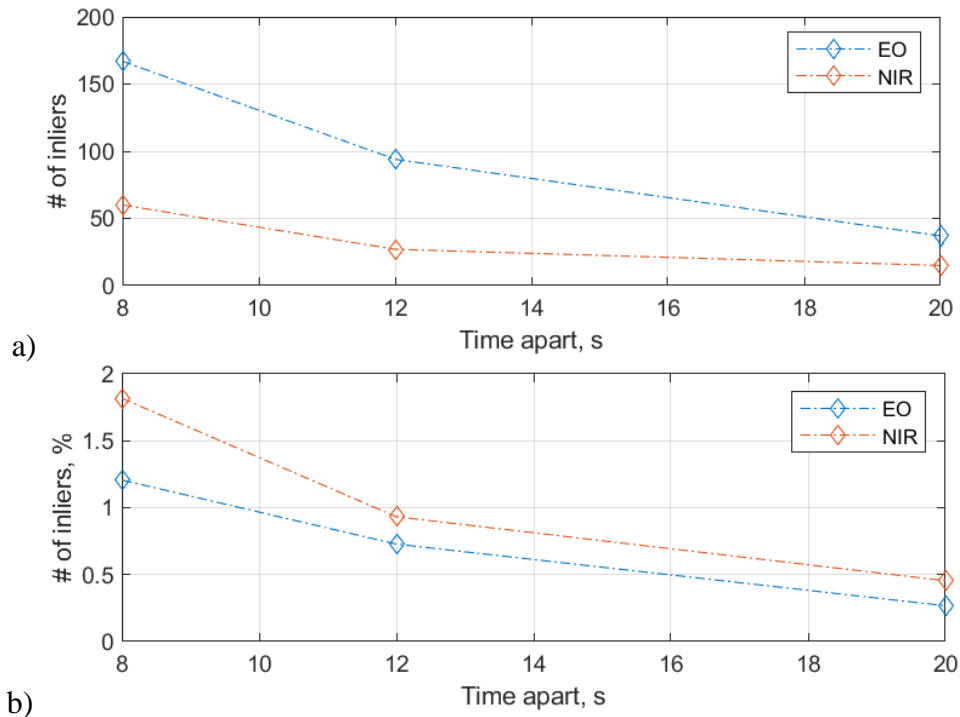


Figure 34. The number of matching features (a) and a percentage of matching features (b) vs sampling rate.

To conclude this discussion, Fig. 36 shows the results of pairwise comparison of the EO and NIR images. The point of this comparison is to show that the EO and NIR images provide different sets of features that may be complementary to each other. Specifically, it shows the comparison of the first and last images in the 14-image series of Fig. 24.



b) Figure 35. The number of inliers (a) and a percentage of inliers (b) vs sampling rate.

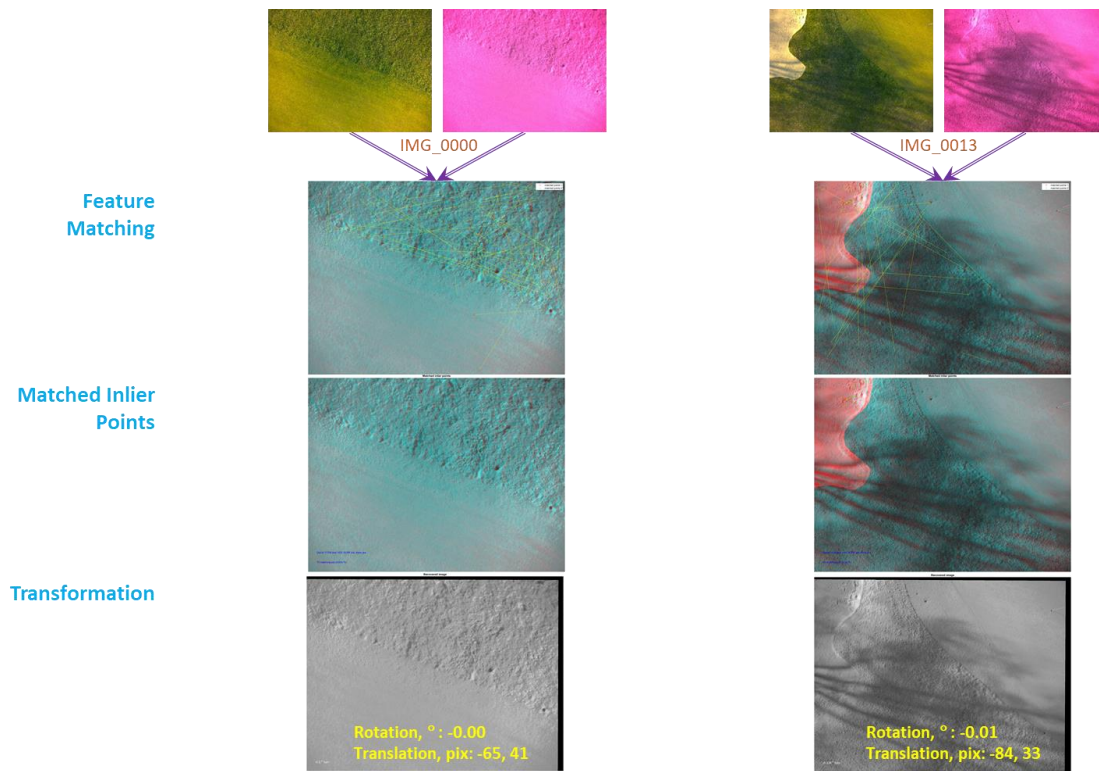


Figure 36. The first and last EO/NIR image matching.

The comparison for the first and last images reveals the following. While 11,706 features were found in the first EO image vs 1,031 features found in the first NIR image, only 73 features (0.6%) seem to match each other. In fact, there were only 7 inliers or 0.06% ($100 * 7 / \max(11706, 1031)$). Similarly, for the last image, 1,919 features were found in the EO image vs 1,441 features found in the NIR image. There were only 45 (2.3%) matched features, and only 13 (0.7%) inliers. Comparing these numbers to what is shown in Figs. 34 and 35, it becomes clear that the same feature-finding method finds mostly different features in the EO and NIR images representing the same scene.

The feature matching and matched inlier stages for the last images are zoomed in in Figs. 37, 38. While Fig. 37 visualizes 45 seemingly matched features (which are all over the place, i.e., not collinear), Fig. 38 shows only 13 inliers (located in the upper left corner of the image).

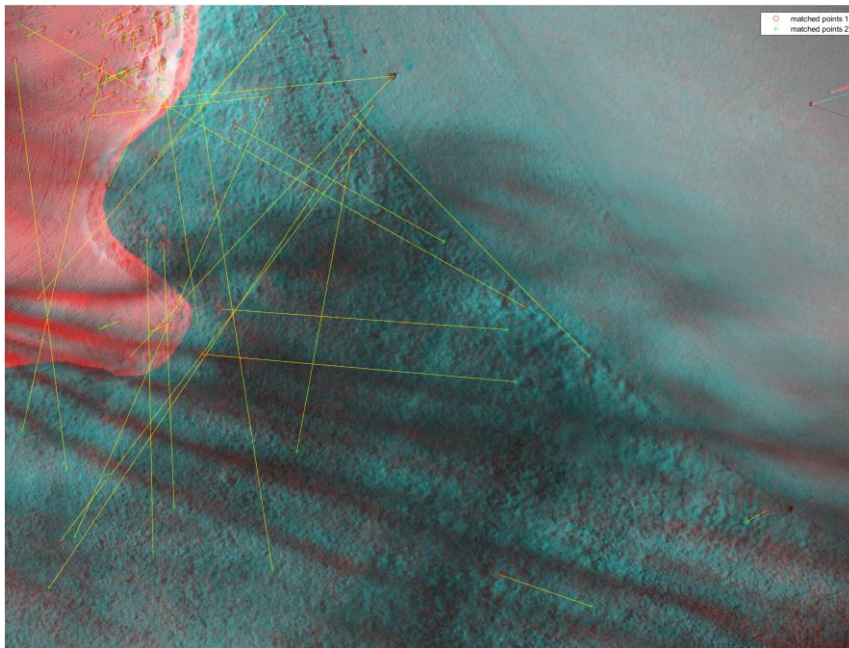


Figure 37. The last EO vs last NIR image matching.

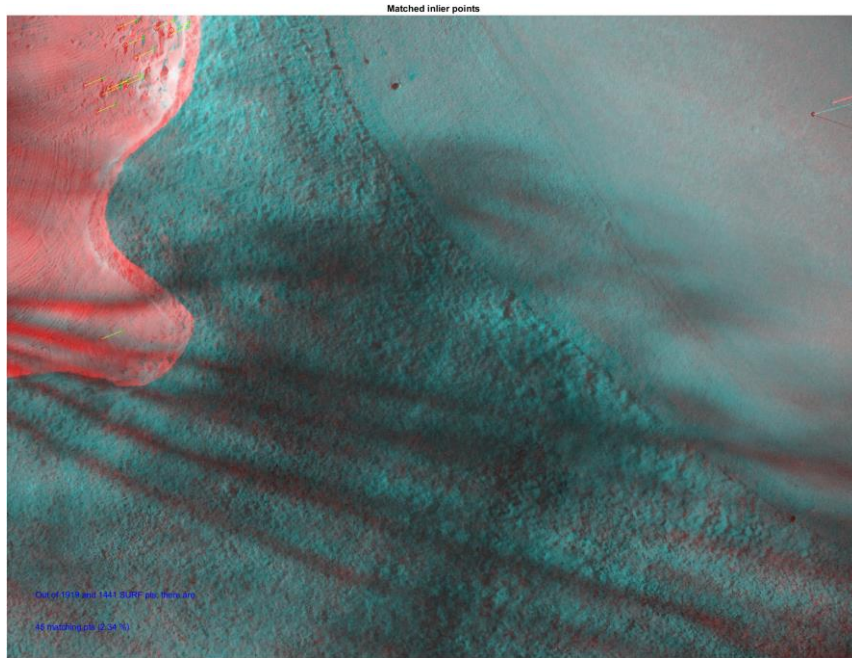


Figure 38. Inliers for the last EO - last NIR image pair.

E. SUMMARY

To summarize:

- Three feature finding methods (out of eight explored), specifically, ORB, FAST and SURF, seem to provide the best feature finding results for both EO and NIR spectra (more matched features found faster) (Fig. 30)
- Features found by EO and NIR sensors, seem to be complimentary (not necessarily the same) (Fig. 36)
- If two consecutive images feature about the same scene, many features can be matched between the images, even though they are far apart (slow flight, high altitude flight, high sampling rate). Even for the specific set of data provided by GreenSight, when the Dreamer UAS was flying very low and images were taken 10s or even 20s apart, there was a healthy number of the matching features and inliers to produce a transform (Figs. 34, 35)
- This is not the case, if the scene changes (like between IMG_0001 and IMG_0002, IMG_0004 and IMG_0005, IMG_0006 and IMG_0007, IMG_0008 and IMG_0009, IMG_0011 and IMG_0012, IMG_0012 and IMG_0013 provided by GreenSight) (Fig. 31)
- As of October 2021, it is unclear what other parameters were recorded by GreenSight Dreamer UAS while flying a trajectory shown in Fig. 25 (including the EO and NIR sensor attitude with respect to the airframe and their zoom setting), whether these flight parameters were synchronized with imagery data (relatively high roll and pitch angles shown in Table 6 seem to contradict imagery data), whether the images were taken by both sensors simultaneously. It is also unknown how GreenSight 3-band sensors were mounted during that flight (whether they had a (joint) gimble mount). The only available data provided in Table 6 did not allow to conduct further analysis



VI. CONCLUSION

The research questions formulated for this study can be answered as follows:

- *Whether using multiple spectral bands has any benefits compared to a standard EO sensor or EO sensor combined with IR sensor? That includes benefits of having a spectral profile of surrounding background area and objects from the standpoint of more reliable/precise DCT.*

Yes, using multiple spectral bands does seem to have benefits compared to a standard EO sensor. In two considered applications (Sections VI and V), using five spectral bands (Blue, Green, Red, Red edge, NIR) of a MicaSense RedEdge-MX sensor (Section IV) or even just two spectral bands (EO and NIR) of a custom GreenSight 3-band MS camera package (Section V), as opposed to just one grayscale image produced from a standard EO sensor output, improvements in reliability (multiple bands complement each other) and accuracy were clearly demonstrated.

- *What are the limitations of using MS sensors and CV/AI algorithms to process data from the standpoint of operating environment, terrain, altitudes, object size and material, time of the day, weather, number of spectral bands, resolution, narrow field of view, addition of a downwelling light sensor)?*

For two applications considered in this study (object detection and vision-aided navigation), the CV/AI algorithms that were developed in the MATLAB development environment worked reliably and did not take more resources than it would be required for processing a standard EO and/or IR sensor output. In a parallel effort, similar algorithms, developed for a standard EO sensor, were successfully deployed on a secondary computer of a Scan Eagle UAS. Due to NDAA'20 ban on flying specific UASs (which included UAS available at NPS and equipped with a MicaSense sensor), and constraints related to coronavirus pandemic, the effects of the operating environment and UAS mission parameters were not addressed.



- *What computational resources would be required to enable DCT capability on board of COTS sUAS?*

The developed algorithms can successfully be run in real time on a secondary computer added to a COTS Group 1 / Group 2 UAS. For example, on a DJI Manifold 2 (currently banned), a high-performance embedded computer, that comes with pre-installed Ubuntu meaning support for Linux, CUDA, OpenCV, and ROS. NPS' ScanEagle UAS takes advantage of using NVIDIA Jetson TX2. As mentioned above, a vision-based GPS-free navigation algorithm developed for a EO sensor, is now under development.

- *Whether an onboard MS sensor and available feature-finding and matching techniques can contribute to enabling GPS-free navigation for aerial vehicles?*

Yes, the MS sensors assessed in this study demonstrated a capability to find thousands of features that can be used for image matching / tracking. While imagery used in this study included a low-altitude flight profile only (10m AGL), the earlier study showed that enough features can be obtained for high-altitude flights (~2km MSL), even in a feature-poor condition, as well [5].

- *What accuracy could be expected when utilizing the non-standard navigational fixes provided by MS sensors?*

This question is yet to be addressed (the current UAS constraints precluded from addressing this research question of this study). A fully-equipped UAS, recording UAS (sensor) position and attitude while taking MS imagery needs to be employed. Using a matching with a satellite imagery technique reported in [6], the positional accuracy while flying a racetrack pattern at ~2 km AGL was estimated to be within 100m. The expectation is that for the low-flying UAS it could be improved significantly (down to 1cm). This question still needed to be addressed in the future research.



The overall conclusion from this study is that utilizing sUAS equipped with MS sensor and CV/AI algorithms may be very beneficial to DoD and DoN offering new and enhancing existing capabilities. As such, the recommendation is to continue the assessment of technologies discussed in this study.



THIS PAGE INTENTIONALLY LEFT BLANK



LIST OF REFERENCES

- [1] EO/IR Surveillance. VectorNAV, www.vectornav.com/applications/eo-ir-surveillance.
- [2] Tingley, B., New Systems For Navigation In GPS Denied Combat Environments Tested In Air Force's Agile Pod, The Drive, June 29, 2021, www.thedrive.com/the-war-zone/41330/new-systems-for-navigation-in-gps-denied-combat-environments-tested-in-air-forces-agile-pod.
- [3] Hoshizaki, T., Andrisani, D., Braun, A., Mulyana, A., Bethel., J., “Optical Navigation Systems,” AIAA Guidance, Navigation, and Control Conference, Austin , TX, August 11-14, 2003, arc.aiaa.org/doi/abs/10.2514/6.2003-5353.
- [4] Chahl, J., and Dawadee, A., “Towards an Optical Aircraft Navigation System,” Applied Mechanics and Materials, vol. 629, Trans Tech Publications, Oct. 2014, pp. 321–326. doi: [10.4028/www.scientific.net/amm.629.321](https://doi.org/10.4028/www.scientific.net/amm.629.321).
- [5] Yakimenko, O., “Assessment of Integrated IMU/EO-Sensor Navigation System for Aerial Vehicles,” NRP Technical Report, Naval Postgraduate School, Monterey, CA, 2020.
- [6] Decker, R., and Yakimenko, O., “On the Development of an Image-Matching Navigation Algorithm for Aerial Vehicles,” Proceedings of the IEEE Aerospace Conference, Big Sky, MT, March 4-11, 2017.
- [7] Multispectral vs Hyperspectral Imagery Explained, GISGeography, June 9, 2021. gisgeography.com/multispectral-vs-hyperspectral-imagery-explained.
- [8] Airborne Hyperspectral and Ground-Truth Technologies, Headwall Photonics, 2016. www.headwallphotonics.com .
- [9] Overview of Agricultural Indices, MicaSense, July 01, 2021, support.micasense.com/hc/en-us/articles/227837307.
- [10] Talmadge, E., “64 Years after Korean War, North Still Digging up Bombs,” AP NEWS, July 24, 2017. <https://apnews.com/dd6256bad51e458cb2e8a1bf64b5c2b6>.
- [11] Liu, H., and Bo, L., “Machine Learning and Deep Learning Methods for Intrusion Detection Systems: A Survey,” Applied Sciences 9(20), 4396, Oct. 2019.
- [12] Lee, W.L., “Assessment of Foreign Object Debris Management Using Group 1 Unmanned Aerial Systems,” M.S. Thesis, NPS, Sep. 2018.
- [13] Dorafshan, S., Thomas, R.J., Coopmans, C., and Maguire, M., “Deep Learning Neural Networks for SUAS-Assisted Structural Inspections: Feasibility and Application,” In the IEEE International Conference on Unmanned Aircraft Systems (ICUAS), Dallas, TX, 2018, pp. 874–882.
- [14] deSmet, T., Nikulin, A., Frazer, W., et al., “Drones and ‘Butterflies’: A Low-Cost UAV System for Rapid Detection and Identification of Unconventional Minefields,” The Journal of Conventional Weapons Destruction, 22(3), 2018.
- [15] Tan, K.H., “Vision-Based Relative Position Estimation and Intercept Trajectory Planning for Small Unmanned Aircraft Systems,” M.S. Thesis, NPS, Sep. 2019.



- [16] Qi, Z., Li, X., Li, H., and Liu, W., “First Results from Drone-Based Transient Electromagnetic Survey to Map and Detect Unexploded Ordnance,” *IEEE Geoscience and Remote Sensing Letters*, 2020, pp. 1–5. doi.org/10.1109/LGRS.2019.2962754.
- [17] Hoonhout, B.M., Radermacher, M., Baart, F., and van der Maaten, L.J.P., “An Automated Method for Semantic Classification of Regions in Coastal Images,” *Coast. Eng.*, 105, 2015, pp. 1–12. <https://doi.org/10.1016/j.coastaleng.2015.07.010>.
- [18] Hermann, D., “Morphodynamic Classification of Coastal Regions Using Deep Learning through Digital Imagery Collection,” M.S. Thesis, NPS, Sep. 2018.
- [19] Tulldahl, H.M., Philipson, P., Kautsky, H., and Wikström, S.A., “Sea Floor Classification with Satellite Data and Airborne LiDAR Bathymetry,” *Ocean Sens. Monit.* V, 8724, 87240B, 2013. <https://doi.org/10.1117/12.2015727>.
- [20] Salamati, N., Larlus, D., Csurka, G., and Süssstrunk, S., “Semantic Image Segmentation Using Visible and Near-Infrared Channels,” *Lect. Notes Comput. Sci.* (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), 7584 LNCS, 2012, pp.461–471. https://doi.org/10.1007/978-3-642-33868-7_46.
- [21] Sensor Comparison, MicaSense. micasense.com/compare-sensors.
- [22] Trinity F90Plus mapping drone, Quantum Systems, Germany, 2021, www.quantum-systems.com/project/trinityf90plus-mapping-drone.
- [23] Announcing the Dreamer Drone with 60 Minutes of Usable Flight Time, GreenSight, 2021 www.greensightag.com/logbook/announcing-the-dreamer-drone-with-60-minutes-of-flight-time.
- [24] Redmon, J., and Farhadi, A., “YOLO9000: Better, Faster, Stronger,” *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, 2017, doi: 10.1109/CVPR.2017.690
- [25] Garg, P., Chowdhury, D.R., and More, V.N., “Traffic Sign Recognition and Classification Using YOLOv2, Faster RCNN and SSD,” *10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Kanpur, India, 2019, pp. 1–5. doi: 10.1109/ICCCNT45670.2019.8944491.
- [26] Zhao, Z.-Q., Zheng, P., Xu, S.-T., and Wu, X., “Object Detection with Deep Learning: A Review,” *IEEE Transactions on Neural Networks and Learning Systems*, 30 (11), 2019, 21.
- [27] Ren, S., He, K., Girshick, R., and Sun, J., “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” 2016, ArXiv:1506.01497 [cs.CV].
- [28] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., and Berg, A.C., “SSD: Single Shot MultiBox Detector,” In *Computer Vision – ECCV 2016*, edited by Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, 9905: 21–37. *Lecture Notes in Computer Science*. Cham, Switzerland: Springer International Publishing, doi: 10.1007/978-3-319-46448-0_2.
- [29] He, K., Zhang, X., Ren, S., and Sun, J., “Deep Residual Learning for Image Recognition,” 2015, pp. 1-12. arXiv:1512.03385 [cs.CV].
- [30] Takumi, K., Watanabe, K., Ha, Q., Tejero-De-Pablos, A., Ushiku, Y., and Tatsuya Harada, “Multi-spectral Object Detection for Autonomous Vehicles.” *Thematic*



Workshops of ACM Multimedia, ACM Press, Mountain View, CA, 2017, pp. 35–43. doi:10.1145/3126686.3126727.

- [31] 31Bodla, Na., Singh, B., Chellappa, R., and Davis, L.S., “Soft-NM —Improving Object Detection with One Line of Code,” 2017, pp. 1-9. ArXiv:1704.04503 [cs.CV].
- [32] Powers, D., “Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation,” Machine Learning Technology, January, 2008, pp. 1-23.
- [33] Stehman, S.V., “Selecting and Interpreting Measures of Thematic Classification Accuracy,” Remote Sensing of Environment 62 (1), 1997, pp.77–89. doi: 10.1016/S0034-4257(97)00083-7.
- [34] Math Works, Inc. MATLAB, version 2020b. Natick, MA: The Math Works, Inc., 2020. www.mathworks.com [Accessed December 28, 2020].
- [35] Everingham, M., Van Gool, L., Williams, C.K.I, Winn, J., and Zisserman, A., “The Pascal Visual Object Classes (VOC) Challenge,” International J. of Computer Vision 88 (2), 2010, pp. 303–38. doi:10.1007/s11263-009-0275-4.
- [36] Rosten, E., and Drummond, T., “Fusing Points and Lines for High Performance Tracking,” Proceedings of the IEEE International Conference on Computer Vision, vol. 2, 2005, pp. 1508–1511.
- [37] Shi, J., and Tomasi, C., “Good Features to Track,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 1994, pp. 593–600.
- [38] Harris, C., and Stephens, M., “A Combined Corner and Edge Detector,” Proceedings of the 4th Alvey Vision Conference, August 1988, pp. 147-151.
- [39] Rublee, E., Rabaud, V., Konolige, K., and Bradski, G., “ORB: An Efficient Alternative to SIFT or SURF,” Proceedings of the 2011 International Conference on Computer Vision. Barcelona, Spain, 2011, pp. 2564–2571.
- [40] Leutenegger, S., Chli, M., and Siegwart, R., “BRISK: Binary Robust Invariant Scalable Keypoints,” Proceedings of the IEEE International Conference on Computer Vision, 2011.
- [41] Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L., “SURF: Speeded Up Robust Features,” Computer Vision and Image Understanding, vol. 110, no.3, 2008, pp. 346–359.
- [42] Alcantarilla, P.F., Bartoli, A., and Davison, A.J., “KAZE Features,” ECCV 2012, Part VI, LNCS 7577, 2012, p. 214.
- [43] Nister, D., and Stewenius, H., “Linear Time Maximally Stable Extremal Regions”, Lecture Notes in Computer Science, Proceedings of the 10th European Conference on Computer Vision, Marseille, France, 2008, pp. 183–196.
- [44] Matas, J., Chum, O., Urba, M., and Pajdla, T., “Robust Wide Baseline Stereo from Maximally Stable Extremal Regions,” Proceedings of the British Machine Vision Conference, 2002, pp. 384-396.
- [45] Obdrzalek D., Basovnik, S., Mach, L., and Mikulik, A., “Detecting Scene Elements Using Maximally Stable Colour Regions, Communications in Computer and Information Science, 82, 2009, pp. 107–115.



- [46] Dalal, N., and Triggs, B., “Histograms of Oriented Gradients for Human Detection”, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol.1, 2005, pp. 886–893.

