

# Multi-Modal Triangulation for In-situ Bore-sight Calibration of Frame and Line Cameras Onboard GNSS/INS-assisted UAVS

Meghdad Hasheminasab<sup>1</sup>, Tian Zhou<sup>1</sup>, Ayman Habib<sup>1\*</sup>

Lyles School of Civil Engineering, Purdue University, 550 Stadium Mall Dr., West Lafayette, IN 47907, USA – (hashemin, zhou732, ahabib)@purdue.edu

## Commission II

**KEY WORDS:** UAVs, Line Cameras, Frame Cameras, Multi-Modal Image Matching, System Calibration, Orthophoto Generation

### ABSTRACT:

Unmanned aerial vehicles (UAVs) equipped with integrated global navigation satellite systems/inertial navigation systems (GNSS/INS) together with RGB and hyperspectral (HS) cameras have become popular data acquisition tool for several applications. To derive accurately georeferenced products from such systems, the spatial and rotational offset between the onboard sensors and the GNSS/INS unit must be determined. While the spatial relationship can be measured manually, establishing the angular offset – i.e., bore-sight angles – is more challenging and requires a calibration strategy. Given that the majority of RGB cameras are based on frame imaging mechanism, bore-sight calibration of these sensors is usually conducted through automated triangulation of overlapping images. On the other hand, most of current HS cameras are based on push-broom technology – also known as line cameras – which capture 1D images at a time. Consequently, automated triangulation of these non-overlapping images is not possible and thus, current bore-sight calibration strategies for HS cameras are based on calibration missions where special targets are deployed in the study site and are manually measured in the imagery. Although bore-sight calibration missions can lead to accurate system calibration parameters, they are expensive and labor intensive activities. To address these limitations and motivated by the new trend of UAVs equipped with both RGB and HS cameras, this study proposes a multi-modal triangulation approach to conduct an in-situ bore-sight calibration for the two cameras, simultaneously. Experimental results over an agricultural field and an urban area show that the proposed approach results in orthophotos with high visual quality and geolocation accuracy.

## 1. INTRODUCTION

Recent advances in remote sensing technologies have made UAVs equipped with global navigation satellite systems/inertial navigation systems (GNSS/INS) together with RGB and hyperspectral (HS) cameras cost-effective and popular for several applications such as digital agriculture (Fu et al., 2021), forestry (Cao et al., 2021), and more recently, urban mapping (Wang et al., 2021). In most of these applications, it is crucial to derive accurately georeferenced products, e.g., point clouds and orthophotos. To do so, system calibration – which includes estimation of internal characteristics of the cameras, known as interior orientation parameters (IOP), as well as system mounting parameters relating the imaging sensors to the onboard GNSS/INS unit – needs to be carried out.

IOP include the principal point coordinates, principal distance and distortion parameters. For consumer-grade RGB cameras these parameters are usually derived through an indoor calibration procedure (He and Habib, 2015). On the other hand, IOP of HS cameras are often provided by the manufacturer and are relatively accurate and stable over time.

Mounting parameters include the spatial and rotational offsets between the camera and GNSS/INS unit. While the spatial offset – i.e., lever-arm – can be manually measured, estimation of rotational components – i.e., bore-sight angles – requires a system calibration procedure. The majority of RGB cameras onboard UAV systems use a frame imaging mechanism. The 2-

dimensional (2D) nature of frame images makes it possible to acquire images with high percentage of overlap and side-lap. Consequently, automated triangulation of overlapping images with substantial number of tie features can be implemented for in-situ bore-sight calibration of RGB cameras. However, in the case of HS cameras, due to the extensive volume of the acquired data, sensors are often designed as push-broom scanners – also known as line cameras. Hence, a scene – which is a 2D coverage of an area on the ground – is generated by concatenating multiple 1D HS images that are captured sequentially. Therefore, no conjugate features can be identified among the scan lines in a given flight line.

The majority of current approaches for bore-sight calibration of UAV-based HS cameras require deployment of ground control points (GCPs) (Zhang et al., 2015). Establishing GCPs is time-consuming and expensive especially in applications such as digital agriculture where frequent calibrations are required throughout the growing season to ensure high accuracy of generated multi-temporal orthophotos. Although in recent years, some strategies have been developed that eliminate the need for deployment of GCPs (Habib et al., 2018), they still require manual measurement of tie points among overlapping scenes.

In order to overcome the abovementioned limitations in current bore-sight calibration techniques for UAV-based HS cameras, and motivated by the new trend of UAVs equipped with both RGB frame and HS line cameras, this paper proposes a multi-modal feature matching and triangulation strategy to conduct an in-situ

\* Corresponding author

boresight calibration for the two cameras simultaneously. More specifically, the main contribution of this study is developing strategies for establishing conjugate features among RGB image and HS scenes and incorporating those features in a unified bundle adjustment (BA) with system self-calibration.

The remainder of this paper is organized as follows: Section 2 introduces data collection platform as well as the datasets used in this study, Section 3 describes the proposed multi-modal triangulation framework, Section 4 presents the experimental results, and Section 5 provides conclusions and recommendations for future work.

## 2. DATA ACQUISITION SYSTEM AND DATASETS DESCRIPTION

In this study, two datasets collected over an agricultural field – hereafter referred to as Agriculture dataset – and an urban area – hereafter denoted as Building dataset – are used to validate the feasibility of the proposed approach. These datasets were acquired using a custom-built UAV, illustrated in Figure 1, that consists of a Trimble APX-15 v3 GNSS/INS unit for georeferencing, a Sony  $\alpha$ 7R III RGB camera, and the Headwall’s visible and near-infrared (VNIR) HS sensor. The GNSS/INS unit has an expected post-processing positional accuracy in the range of  $\pm 2$ -5 cm, and attitude accuracy of  $\pm 0.025$  and  $\pm 0.08$  for the roll/pitch and heading, respectively. The RGB camera has a  $7952 \times 5304$  sensor array size with  $4.5 \mu\text{m}$  pixel pitch and a lens with 35mm nominal focal length. The VNIR line camera covers 270-273 spectral bands ranging between 398 and 1000 nm. The scan line consists of 640 pixels with a detector pitch of  $7.5 \mu\text{m}$  and a lens with 8.2 mm focal length.



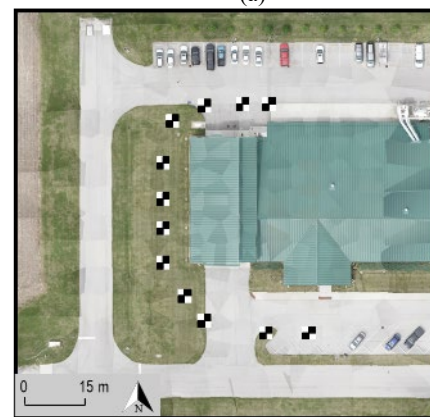
**Figure 1.** The UAV-based mobile mapping system and onboard sensors used in this study.

The two sites used in this study are located at Purdue’s Agronomy Center for Research and Education (ACRE). The Agriculture dataset was captured 50 days after sowing (DAS) over sorghum plants with an average crop height of 1.2 m. On the other hand, the Building dataset, was captured over an area with different geometric features, i.e., building roof, pavement, and

grass. To assess the absolute accuracy derived from the proposed approach, 14 and 12 checkerboard targets were deployed in the agricultural field and urban area, respectively. The ground coordinates of these targets were determined using a real-time kinematic (RTK) GNSS survey with a Trimble R10 GNSS receiver. Figure 2 shows sample orthophotos of the two study sites along with an enhanced representation of the ground targets. The data collection dates, flying height, ground sampling distance (GSD), and percentage of overlap/side-lap among images are reported in Table 1. As presented in this table, the GSD of the frame camera – i.e., 0.5 cm at 44 m flying height – is almost 8 times smaller than that for the line camera – i.e., 4 cm at 44 m flying height. This difference in spatial resolution of the two cameras is illustrated in Figure 3 with the help of overlapping RGB image/HS scene captured by the frame/line cameras over the agricultural field.



(a)

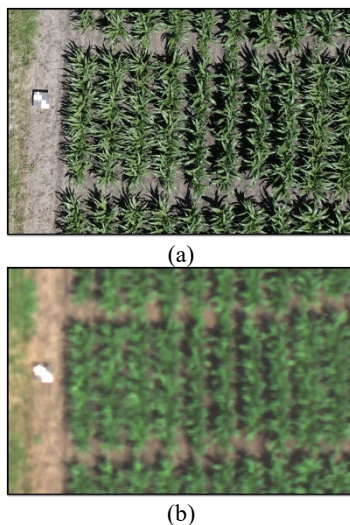


(b)

**Figure 2.** Sample orthophotos of the study sites along with an enhanced representation of the ground targets: (a) Agriculture field, and (b) Building datasets.

Dataset	Acquisition Date	Camera	Flying height (m)	GSD (cm)	Overlap/Side-lap (%)
Agriculture	2020/07/02	Frame	44	0.5	80/75
		Line		4	NA/50
Building	2021/05/27	Frame	44	0.5	80/84
		Line		4	NA/50

**Table 1.** Data collection dates, flying height, GSD, and overlap/side-lap for the two datasets used in this study.



**Figure 3.** Sample overlapping RGB image and HS scene captured at the same height (44 m): (a) RGB frame image with 0.5 cm GSD, and (b) HS line camera scene with 4 cm GSD.

### 3. METHODOLOGY

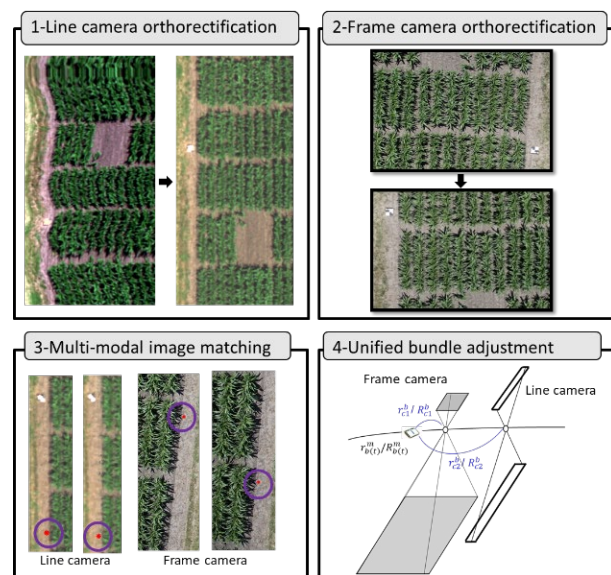
In this section, the proposed multi-modal image matching for simultaneous boresight calibration of frame and line cameras onboard UAVs is presented. The proposed framework is implemented in four steps, as outlined in Figure 4. As shown in the first block of Figure 4, due to turbulence in the UAV trajectory, concatenating sequential scans in push-broom scanners results in wavy patterns in the raw scenes. This will impede the image matching process. Hence, in the first step of the proposed strategy, the raw HS scenes are partially ortho-rectified. The term “partially” refers to the fact that only a rough digital surface model (DSM) and nominal system calibration parameters are used in this process. In the rectification procedure, a look-up table is created to store the raw image coordinates (which are needed for the BA procedure) for each cell of the ortho-rectified scene.

As mentioned earlier, RGB frame images are of much higher spatial resolution (~8 times in this study) compared to HS scenes when captured at the same flying height. Such significant scale difference makes the same feature look different in the RGB and HS imagery. Therefore, existing scale-invariant image matching algorithms such as SIFT (Low 2004) and SURF (Bay et al., 2006) would result in high percentage of outliers when it comes to processing UAV-images with such scale differences. To mitigate this problem, in the second step, RGB images are down-sampled to the same GSD of HS scenes in the partial ortho-rectification process. Similar to the first step, a look-up table is created to store the raw image coordinates for ortho-rectified cells. One should note that first and second steps can be carried out in parallel.

Next, SIFT algorithm is applied on all partially ortho-rectified images/scenes. In this step, geolocation information of the ortho-rectified images/scenes is used to reduce the matching search space by conducting a window-based image matching strategy (Hasheminasab et al., 2021). As a result of this step, conjugate points between overlapping RGB images, HS scenes, and RGB image/HS scenes are identified.

Finally, in step 4, a unified bundle adjustment engine based on the collinearity equations is implemented for simultaneous triangulation of frame and line camera imagery. More specifically, two collinearity equations are established for each

image point identified in the previous step. The unknown parameters include the frame and line cameras’ boresight angles, and ground coordinates of tie points. Also, the position and orientation of the GNSS/INS trajectory at the time of frame/line camera exposures are adjusted through incorporating them in the BA as pseudo observations – i.e., direct observations of the unknowns. The BA process is conducted iteratively where in each iteration tie points with large back-projection residuals are detected and removed as potential outliers. The procedure is repeated until no points with residuals larger than a predefined threshold remain or a maximum number of iterations is reached.



**Figure 4.** Schematic diagram of the involved steps in the proposed multi-modal triangulation framework.

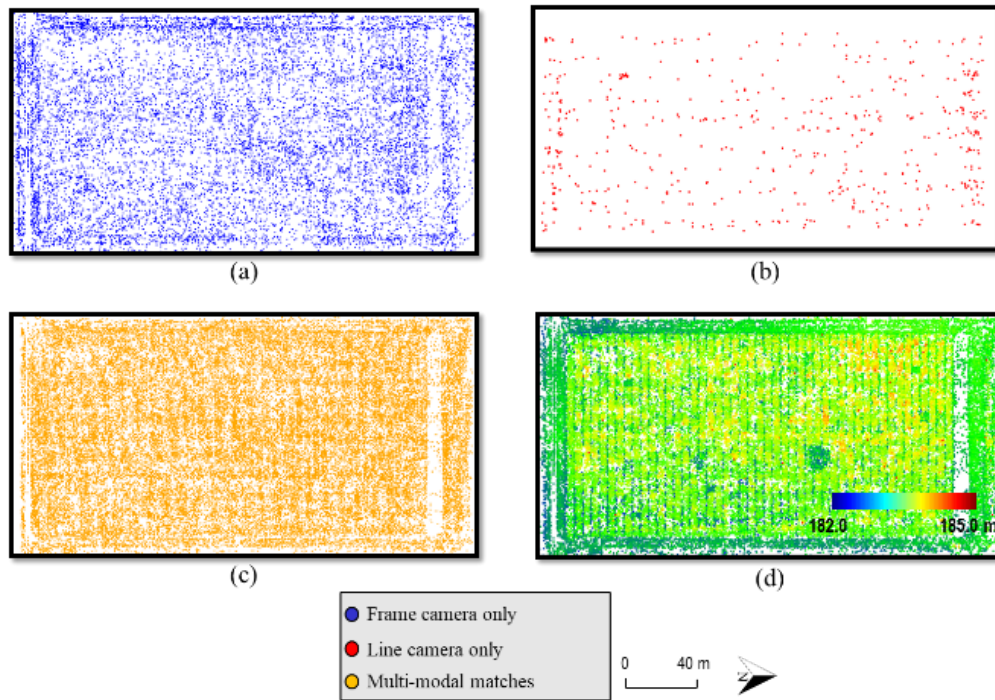
### 4. RESULTS AND DISCUSSION

In this section, the experimental results for the frame and line camera simultaneous in-situ boresight calibration are presented. The experimental results are based on three main objectives as listed below:

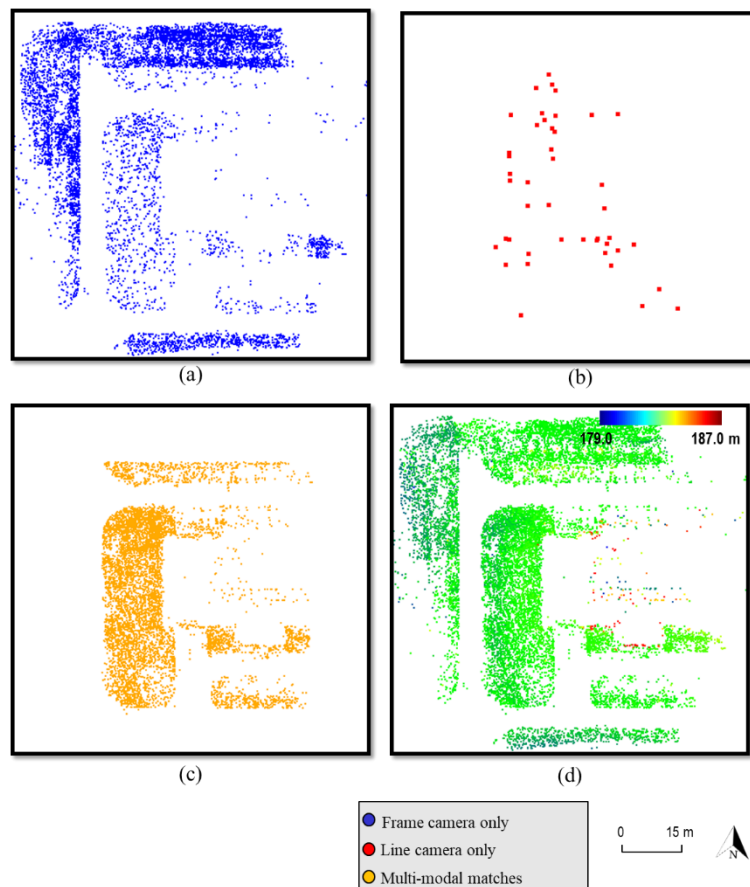
- 1) Evaluation of the proposed approach for identifying well-distrusted and sufficient number of conjugate features between overlapping RGB images and HS scenes.
- 2) Evaluation of bundle adjustment results and absolute accuracy of the reconstruction using the estimated boresight angles.
- 3) Qualitative assessment of the RGB and HS orthophotos generated from the nominal and estimated boresight angles.

#### 4.1 Objective 1 – Evaluation of the Multi-modal Matching for Identifying well-distrusted and Sufficient Number of Conjugate Features

First objective of the experimental results is evaluated through illustration of the reconstructed object points from the multi-modal triangulation with the built-in outlier removal strategy. Figures 5 and 6 show the 3D reconstruction results for the Agriculture and Building datasets, respectively. In these figures, object points that are identified only in frame images are shown in blue, points that are only established between neighbouring line camera scenes are in red, and multi-modal matching results – i.e., points identified in both frame and line camera imagery – are shown in orange. Also, the entire 3D point cloud (colored by height) for the two datasets are illustrated in Figures 5d and 6d.



**Figure 5.** Top view of the 3D reconstruction results for the Agriculture dataset: (a) ~16,000 points identified in frame camera imagery only, (b) ~500 points identified in line camera scenes only, (c) ~49,000 points identified in frame and line camera imagery, and (d) all ~65,500 reconstructed points colored by height.



**Figure 6.** Top view of the 3D reconstruction results for the Building dataset: (a) ~7,800 points identified in frame camera imagery only, (b) 44 points identified in line camera scenes only, (c) ~7,500 points identified in frame and line camera imagery, and (d) all ~15,300 reconstructed points colored by height.

As can be seen in Figures 5 and 6, one can note that a high percentage of object points are established among frame RGB and HS imagery, which is an indication of a successfully conducted multi-modal matching for both datasets. This percentage is ~75% and ~50% for the Agriculture and Building datasets, respectively. The reason behind having less percentage of multi-modal matches in the Building dataset is that line camera scenes covered a smaller area on the ground compared to RGB images for this dataset. This can be visually observed by comparing Figures 6a and 6c. It can also be noted that very few conjugate points (less than 0.1% for both datasets) are only identified in neighbouring line camera scenes. This confirms the hypothesis that using line camera scenes alone is not sufficient for deriving reliable conjugate points that can be used in the BA with system self-calibration.

#### 4.2 Objective 2 – Quantitative Evaluation of the Estimated Boresight Angles

In this subsection, bundle adjustment results and reconstruction accuracy of the proposed strategy is presented. The square root of a-posteriori factor ( $\hat{\sigma}_0$ ) resulting from the unified BA was 2.1 and 1.4 (in pixel units) for the Agriculture and Building datasets, respectively. Such small value of  $\hat{\sigma}_0$  is an indication of a high quality of fit between observations and estimates of unknown parameters as described by the collinearity equations. A slightly smaller value of  $\hat{\sigma}_0$  for the Building dataset can be attributed to a higher percentage of matching inliers due to availability of more distinctive features in this study site. Table 2 reports the nominal and estimated boresight angles along with their standard deviations for the two datasets. As per the reported values in this table, one can observe that there is a significant difference between the nominal and estimated angles for both line and frame cameras and between the two datasets. As an example, 1.2° variations in  $\Delta\phi$  (see nominal and estimated values for the line camera in Agriculture dataset) can cause a displacement in the object space with a magnitude of ~92 cm (~23 times the GSD of line camera) when flying at 44 m. Also, small standard deviation values (in the range of 0.6 to 1.8 arcminutes) is an indication of a precise estimation of the boresight angles using the proposed multi-modal triangulation framework.

Dataset	Camera	Boresight angles	$\Delta\omega(^{\circ})$	$\Delta\phi(^{\circ})$	$\Delta\kappa(^{\circ})$
Agriculture	Line	Nominal	0.00	0.00	0.00
		Estimated	0.38 ±0.01	-1.20 ±0.01	-0.44 ±0.02
	Frame	Nominal	180.00	0.00	-90.00
		Estimated	179.80 ±0.01	0.88 ±0.01	-89.96 ±0.01
Building	Line	Estimated	0.65 ±0.01	-0.96 ±0.02	-0.20 ±0.03
	Frame	Estimated	179.49 ±0.01	0.60 ±0.02	-90.20 ±0.01

**Table 2.** The nominal and estimated boresight angles along with their standard deviations for the two datasets.

In order to evaluate the absolute accuracy of the estimated boresight angles from the multi-modal triangulation process, first the center of the checkerboard targets are manually measured in the RGB/HS images. Then the object coordinates of these targets are estimated through a multi-light ray intersection while using the estimated boresight angles and refined trajectory derived from the BA process. When nominal boresight angles are used, original trajectory is employed in the intersection process.

Table 3 presents the root-mean-square error (RMSE) values of differences between the intersection-derived and surveyed coordinates of ground targets for the two datasets. The reported values reveal that the proposed approach can significantly improve the absolute accuracy when compared to the one derived from the nominal boresight angles. More specifically, the multimodal triangulation leads to horizontal accuracy in the range of 2-3 GSD of the original image for both cameras. A higher vertical RMSE (5-7 times of the GSD) in line camera can be attributed to the weak intersection geometry of this imaging system (i.e., smaller intersection angle).

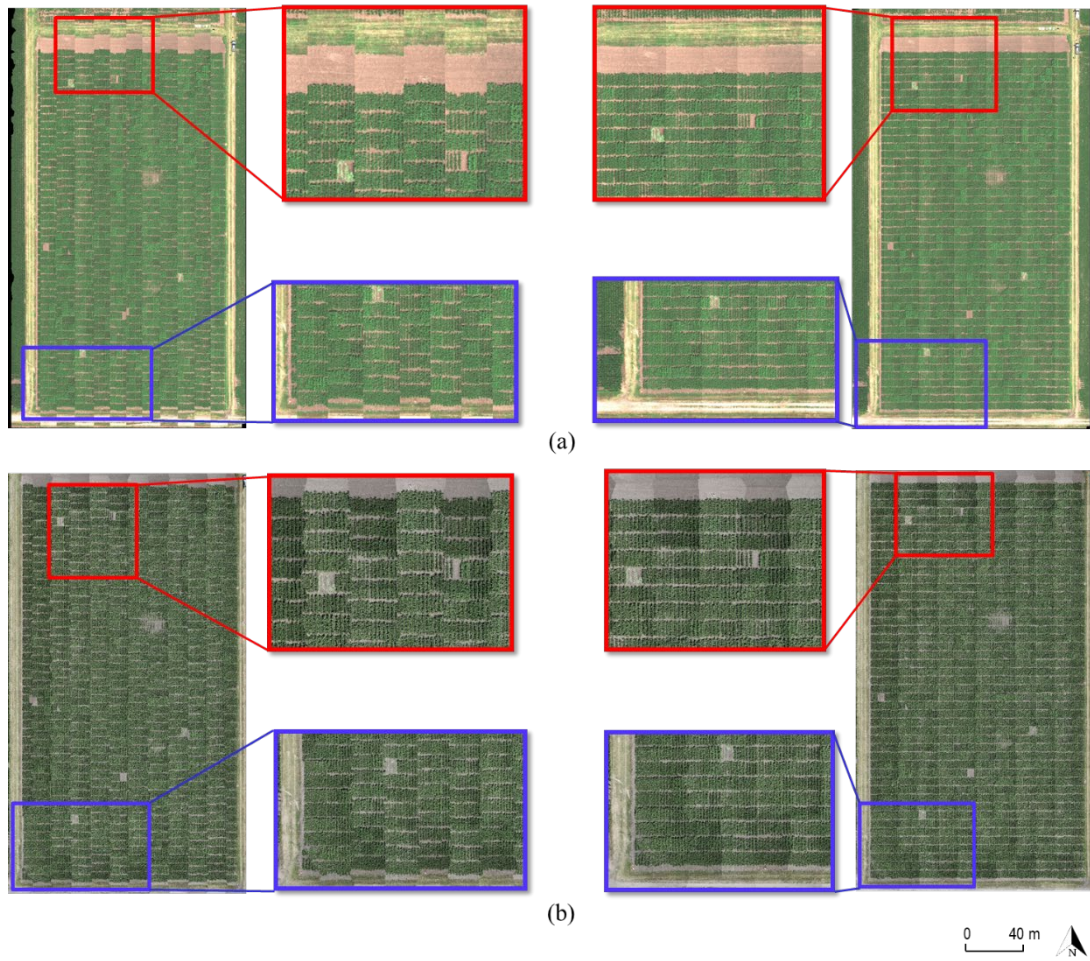
Dataset	Camera	Boresight angles	RMSE X (m)	RMSE Y (m)	RMSE Z (m)
Agriculture	Line	Nominal	0.27	0.15	3.92
		Estimated	0.03	0.03	0.27
	Frame	Nominal	0.04	0.10	0.46
		Estimated	0.03	0.01	0.03
Building	Line	Nominal	0.32	0.20	2.87
		Estimated	0.04	0.03	0.21
	Frame	Nominal	0.04	0.05	0.36
		Estimated	0.01	0.02	0.03

**Table 3.** Absolute accuracy of the multi-modal triangulation procedure for the two datasets used in this study.

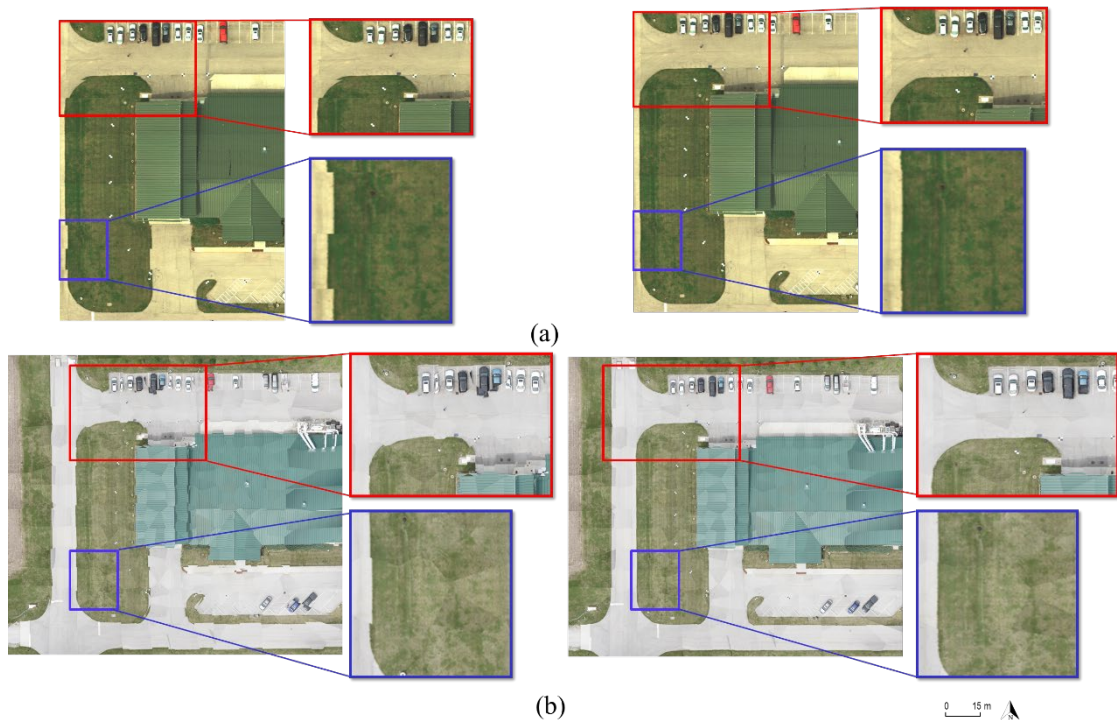
#### 4.3 Objective 3 – Qualitative Evaluation of the Multi-Modal Triangulation Approach

In the last section of the experimental results, the accuracy of the estimated boresight angles is qualitatively assessed by visually inspecting generated orthophotos using the multi-modal triangulation results. Figures 7 and 8 show the RGB and HS orthophotos generated while using nominal and estimated boresight angles for the two datasets. While obvious misalignments are visible in the derived orthophotos when using the nominal boresight angles, we can see that those misalignments have been eliminated when using the boresight angles derived from the proposed strategy.

In summary, from the presented quantitative and qualitative results, one can conclude that the proposed multi-modal triangulation procedure is capable of conducting accurate in-situ boresight calibration for both frame and line cameras.



**Figure 7.** Orthophotos generated from nominal (left) and estimated boresight angles (right) for the Agriculture dataset: (a) HS orthophotos and (b) RGB orthophotos.



**Figure 8.** Orthophotos generated from nominal (left) and estimated boresight angles (right) for the Building dataset: (a) HS orthophotos and (b) RGB orthophotos.

## 5. CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK

In this study, a multi-modal triangulation framework for simultaneous in-situ calibration of frame and line cameras onboard GNSS/INS-assisted UAVs has been introduced. The key motivation for such development is eliminating the need for expensive and time-consuming deployment of ground control points for boresight calibration missions. The proposed strategy consists of four main steps: partial ortho-rectification of line camera scenes for removing wavy patterns in the concatenated HS images, down-sampling of frame images to the same GSD of HS scenes through partial ortho-rectification, window-based feature matching, and unified BA with system self-calibration. The experimental results over an agricultural field and an urban area indicated that the proposed strategy is capable of deriving boresight angles that can lead to accurate 3D reconstruction – i.e., horizontal accuracy in the range of 2-3 times the GSD of original images and vertical accuracy in the range of 2-7 times the GSD – for both frame and line cameras. Also, generated RGB and HS orthophotos from the proposed approach showed a good alignment when compared to those generated from nominal boresight angles.

The agricultural field used in this was in early-growth stage with relatively short plants (in the range of 1-1.5 m). The performance of the proposed approach when dealing with mid-to-late season datasets with tall plants will be investigated in future research.

## ACKNOWLEDGMENTS

The information, data, or work presented herein was funded by the Advanced Research Projects Agency-Energy (ARPA-E), U.S. Department of Energy, under Award Number DE AR0001135. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof. The authors would like to thank Taojun Wang and Karoll Quijano for their work on data acquisition and processing.

## REFERENCES

- Bay, H., Tuytelaars, T., & Gool, L. V. (2006, May). Surf: Speeded up robust features. In *European conference on computer vision* (pp. 404-417). Springer, Berlin, Heidelberg.
- Cao, J., Liu, K., Zhuo, L., Liu, L., Zhu, Y., & Peng, L. (2021). Combining UAV-based hyperspectral and LiDAR data for mangrove species classification using the rotation forest algorithm. *International Journal of Applied Earth Observation and Geoinformation*, 102, 102414.
- Fu, Y., Yang, G., Song, X., Li, Z., Xu, X., Feng, H., & Zhao, C. (2021). Improved estimation of winter wheat aboveground biomass using multiscale textures extracted from UAV-based digital images and hyperspectral feature analysis. *Remote Sensing*, 13(4), 581.
- Habib, A., Zhou, T., Masjedi, A., Zhang, Z., Flatt, J. E., & Crawford, M. (2018). Boresight calibration of GNSS/INS-assisted push-broom hyperspectral scanners on UAV platforms. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(5), 1734-1749.
- Hasheminasab, S. M., Zhou, T., LaForest, L. M., & Habib, A. (2021). Multiscale Image Matching for Automated Calibration of UAV-Based Frame and Line Camera Systems. *IEEE Journal of*

*Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 3133-3150.

He, F., & Habib, A. (2015, May). Target-based and Feature-based Calibration of Low-cost Digital Cameras with Large Field-of-view. In *Proceedings of the ASPRS 2015 Annual Conference, Tampa, FL, USA* (pp. 4-8).

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2), 91-110.

Wang, Z., Wei, L., He, C., & Lu, Q. (2021, July). Ammonia Nitrogen Monitoring of Urban Rivers with UAV-Borne Hyperspectral Remote Sensing Imagery. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS* (pp. 3713-3716). IEEE.

Zhang, A., Hu, S., Meng, X., Yang, L., & Li, H. (2015). Toward high altitude airship ground-based boresight calibration of hyperspectral pushbroom imaging sensors. *Remote Sensing*, 7(12), 17297-17311.