Multiscale Image Matching for Automated Calibration of UAV-Based Frame and Line Camera Systems

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

Abstract—Unmanned aerial vehicles (UAVs) equipped with integrated global navigation satellite systems/inertial navigation systems together with frame and/or line cameras are used for a variety of applications. Geometric system calibration is crucial for delivering accurate products from UAV-based imaging systems. This article presents automated geometric calibration strategies for UAV-based frame and line camera systems to estimate accurate system calibration parameters without the need for ground control points or manual measurements of tie points. The matching strategy used in this article to establish conjugate features among overlapping frame camera images is based on a traditional Structure from Motion technique augmented with several layers of matching outlier removal. On the other hand, a new strategy relying on orthorectified images is introduced for automated feature matching in line camera scenes. Then, a general bundle adjustment procedure with system calibration capabilities for frame and line cameras is presented, where the derived automated tie points are used for estimating accurate geometric system calibration parameters. The proposed approach is evaluated using four datasets—two datasets captured by frame cameras and two datasets captured by line cameras. The results show that the developed automated calibration strategy is capable of producing the same level of absolute accuracy when compared to using manually measured tie points for both frame camera and line camera systems. Results also indicate that the presented automated system calibration approach can be applied to systems even with significant deviation of actual system parameters from their nominal values, and still produce accurate estimates of calibration parameters.

Index Terms—Hyperspectral line cameras, image matching, integrated global navigation satellite system/inertial navigation system (GNSS/INS), RGB frame cameras, system calibration, unmanned aerial vehicles (UAVs).

I. INTRODUCTION

UNMANNED aerial vehicles (UAVs) equipped with integrated global navigation satellite systems/inertial navigation systems (GNSS/INS) and imaging systems are used in a variety of applications for accurate 3-D reconstruction. UAVs provide ease of deployment and variety of choices for imaging and GNSS/INS sensors [1]. Among potential applications, the use of UAVs equipped with RGB and hyperspectral imaging systems in precision agriculture [2]–[5] has expanded rapidly in recent years due to their relatively low cost and ability to provide data with high spectral, spatial, and temporal resolutions [6]. However, to produce accurate 3-D data, system calibration is needed for consumer-grade UAV imaging systems. In this article, “consumer-grade” refers to UAV systems that are equipped with relatively low-cost GNSS/INS units and imaging systems that require system calibration by the user instead of by the manufacturer or other high-end laboratory settings.

Other than estimating internal characteristics of the different sensors, geometric system calibration includes spatial calibration—i.e., determining mounting parameters relating the GNSS/INS unit and camera coordinate systems—as well as temporal calibration—i.e., estimating time delay between the GNSS/INS recorded event markers and actual time of image exposure. Since geometric system calibration is vital to achieving higher levels of accuracy, it has been a major focus in existing research. Spatial and temporal system calibration has been performed on a variety of frame camera imaging sensors with many requiring ground control points (GCPs) to achieve the desired accuracy [7], [8]. Establishing such ground targets is expensive and labor intensive, and more importantly, the distribution and number of GCPs are usually less than optimal to provide adequate control for determining system calibration parameters. To overcome this limitation, there has also been recent research focusing on UAV-based RGB camera [9] and hyperspectral scanner system calibration [10] without the need for GCPs. Yet, even without using GCPs, those techniques have only been applied using manually measured tie points in overlapping images. Manual tie point measurements are known to be accurate but are very time consuming and labor intensive. One should note that although several studies focusing on automated identification of tie features among overlapping images have been introduced in the literature, the majority of these approaches target applications such as image registration, object detection, and multiview 3-D reconstruction. To the best of the authors’ knowledge, automated feature extraction and matching for UAV-based system calibration has not been considered by existing research.
Due to the excessive volume of collected data by hyperspectral cameras, the far majority of these sensors adopt a single line in their focal plane (i.e., they are line cameras). On the other hand, the majority of RGB and thermal cameras onboard UAV systems use a frame imaging mechanism. Frame and line camera systems capture imagery through different data acquisition mechanisms as depicted in Fig. 1. The main difference between frame and line cameras is that a frame camera scene is captured in one instance by a single image, whereas a line camera scene is built by concatenating multiple images captured sequentially. Here, a scene is considered as a 2-D coverage of an area on the ground, while an image represents the output of a single exposure of the light-sensitive elements (pixels) in the image plane. Given this geometry, frame camera’s image coordinates will consist of both $x$ and $y$ values that are variable depending on the image point location which is bounded by the sensor’s angular field of view (AFOV). However, line camera’s $y$ image coordinates will have variable values but $x$ will always be constant depending on the sensor placement in the image plane (for the line camera, the $x$-axis of the image coordinate system is aligned along the flight direction). These acquisition mechanisms cause the automated tie point generation and system calibration process to differ for frame and line camera systems.

For a robust estimation of system calibration parameters with minimum correlation between such parameters, specific flight missions are required. As discussed in [9], to conduct spatial and temporal calibration, it is recommended to derive the system parameters using opposite flying directions at different flying heights, as well as having a variation in the linear and angular velocities. In addition to potential differences among overlapping UAV images such as changes in illumination, viewpoint, and distortions, data acquisition from multiple flying heights results in imagery with varying scale. Such factors adversely affect automated feature extraction and matching in frame and line camera imagery. Fig. 2 illustrates raw imagery acquired from two different flying heights, i.e., 20 and 40 m, for both frame and line cameras. As shown in this figure, conjugate features look different in overlapping images/scenes with significant scale difference.

Although several scale invariant image matching algorithms such as scale invariant feature transform (SIFT) [11] and speeded-up robust features (SURF) [12] have been developed, when dealing with UAV-based images with significant scale differences, these algorithms would result in high percentage of outliers. In order to illustrate the impact of scale difference on the automated matching process, three different stereo-pairs acquired by a frame camera from 1) 40-m flying height, 2) 20-m flying height, and 3) 40 and 20-m flying heights are considered for image matching. The selected stereo-pairs are shown in Fig. 3. Table I shows the matching results for the three different stereo-pairs. In this table, initial matches are based on the SIFT algorithm, with the number of inliers derived through a random sample consensus (RANSAC)-based outlier removal [13]. As reported in Table I, the number of initial correspondences identified in images with different scale is significantly smaller than that for images at the same scale. In addition, while only 14%–18% of initial SIFT matches are detected as outliers for stereo-pairs with the same scale, this value jumps to 37% in case of using images captured from different flying heights. Incorporating such matching results in the bundle adjustment (BA) with self-calibration process can
lead to inaccurate system parameters. Hence, to increase the robustness of the reconstructed object space when dealing with image datasets exhibiting significant scale differences, removing matching outliers is a crucial step.

In addition to the challenges arising from scale differences in imagery, in line camera systems, due to turbulence in the UAV trajectory, concatenating sequential scan lines [as shown in Fig. 1(b)] results in wavy patterns in the generated scenes, which further complicates the process of automated image matching. One should note that although a gimbal mount could mitigate such wavy patterns in line camera scenes, it is not capable of handling large payload such as the coaligned cameras that will be used in this research. Fig. 4 shows two samples of overlapping line camera scenes captured from the same flying height (40 m) but in two different flying passes, along with the corresponding zoomed-in regions. As can be seen in Fig. 4, raw scenes exhibit a wavy pattern and consequently, corresponding features look quite different [see Fig. 4(c) and (d)]. Hence, for reliable extraction of conjugate features in line camera scenes, the impact of the wavy pattern needs to be considered.

This article focuses on fully automated spatial and temporal calibration of frame and line cameras onboard GNSS/INS-assisted UAV imaging systems without the need for GCPs while using a general BA framework. First, a review of existing work dealing with spatial and temporal calibration for frame and line camera systems is presented. Next, the systems specifications and dataset description are provided. The methodology starts by describing the Structure from Motion (SfM) approach used for automated feature extraction and matching for frame cameras while focusing on the matching outlier removal strategies. Next, a new automated feature matching strategy for line camera systems is discussed which applies the SIFT algorithm on partially ortho-rectified images instead of raw images. Here, the term “partially,” refers to the fact that only a rough digital surface model (DSM) as well as nominal system calibration parameters are used in this procedure. Then, the general mathematical model used for the GNSS/INS-assisted BA with system self-calibration process for both frame and line cameras is introduced. The experimental results for frame and line cameras onboard UAV systems are then presented where the ability of the proposed approaches to identify well-distributed, sufficient number of tie features is evaluated. Through these experiments, the resulting system calibration parameters are investigated using four UAV datasets. Last, conclusions and recommendations for future work are discussed.

II. RELATED WORK

Spatial and temporal system calibration is essential for accurate 3-D reconstruction from consumer-grade UAV imaging systems leading to much research over the years. Spatial calibration has been extensively covered in past research [14]–[18]. However, temporal calibration has been addressed to a lesser extent. Chiang et al. [19] proposed a two-step geometric calibration method for a fixed-wing UAV-based imaging system. First, exterior orientation parameters (EOP) were obtained through indirect georeferencing with the help of GCPs. Then, the differences between such EOP and the interpolated GNSS/INS position and orientation of the inertial measurement unit (IMU) body frame
were derived. Those differences were then used in a calibration algorithm to solve for lever arm components, boresight angles, and time delay. Blazaquez [7] introduced a one-step approach for spatio-temporal calibration of manned aircraft multisensor systems. This approach focused on modifying the sensor model to include a time synchronization parameter while using specific flight and GCP configurations. Rehak and Skaloud [8] addressed the system calibration for consumer-grade cameras on micro aerial vehicles. In their proposed approach, lever arm components were derived using a static field calibration introduced in [20], and then camera interior orientation parameters (IOP) and boresight angles were estimated in a self-calibration procedure using GCPs. Next, given the estimated camera IOP and spatial system calibration parameters, the authors modified the mathematical model to include time delay as a parameter in the BA procedure. Yet, all the above system calibration techniques required GCPs for acceptable absolute accuracy and focused on frame camera systems. LaForest et al. [9] presented two approaches—direct and indirect—for estimating geometric calibration parameters for UAV-based frame camera systems without the need for GCPs. The direct approach modified existing collinearity equations to incorporate time delay as a system parameter along with lever arm components and boresight angles while using manually derived tie points. On the other hand, in order to use existing bundle block adjustment packages, the indirect approach exploited a bias impact analysis for the system parameters and their correlations to first solve for time delay, and then estimate spatial calibration parameters. Even though several studies have addressed spatial and temporal calibration of frame cameras, line camera system calibration has primarily focused on spatial calibration. Kocaman et al. [21] focused on modifying the adjustable parameters for self-calibration triangulation of airborne linear array cameras. Habib et al. [10] performed rigorous boresight calibration for a hyperspectral line camera equipped with a GNSS/INS unit onboard a UAV.

System calibration using a bundle block adjustment requires image coordinates measurements of tie points, among other potential observations and unknowns. Manually established tie points are known to be accurate but very time consuming. To address this challenge, some studies have focused on self-calibration while reducing or eliminating the need for manual measurements. Habib et al. [14], [22] focused on LiDAR and frame camera self-calibration using manually derived straight lines and planar patches. Recent research has also focused on SfM techniques for automated in-situ estimation of camera internal parameters. Gneeniss et al. [23] developed an in-flight camera IOP refinement procedure using a least squares surface matching algorithm to align SfM-derived point clouds to LiDAR control points. In a similar study, Zhou et al. [24] used photogrammetric points as well as LiDAR data for refining camera parameters without the need for GCPs. However, there is still a gap when it comes to automated spatial and temporal calibration for frame and line camera systems without the need for GCPs. Moreover, to the best of the authors’ knowledge, there is no existing work that has focused on a unified BA engine for aerial triangulation of frame and line camera systems. This article focuses on a general, fully automated BA with self-calibration including spatial and temporal system parameters for frame and line camera systems without the need for having GCPs.

III. SYSTEM SPECIFICATIONS AND DATASET DESCRIPTION

Feasibility of the proposed research is validated using datasets collected by two UAV systems; a Da-Jiang Innovations (DJI) M200 [25] and a DJI M600 Pro [26]. Both systems relied on the Applanix APX-15 UAV v2 GNSS/INS unit for direct georeferencing with a predicted positional accuracy of 2–5 cm and heading and roll/pitch accuracy of 0.080° and 0.025°, respectively [27]. The M200 platform carried a coalign FLIR Duo Pro R 640 RGB and thermal frame cameras. The RGB sensor array size was 4000 × 3000, with a pixel size of 1.85 μm and a nominal focal length of 8 mm and had an AFOV of 56° [28]. The Uncooled VOx Microbolometer thermal sensor array was 640 × 512 with a pixel size of 17 μm and had a nominal focal length of 19 mm and an AFOV of 32°. Fig. 5 shows the FLIR Duo Pro R and APX-15 configuration onboard the M200 UAV, and illustrates the coordinate systems for the IMU body and cameras frames. The FLIR Duo Pro R utilized a mobile-phone-based app to set camera parameters via Bluetooth. The mobile-phone-based app provides the ability to set the capture interval for the camera and to start and stop triggering. Event feedback to the APX was provided directly by the FLIR Duo Pro R using the “Frame Sync” option. This option outputs a low-voltage transistor to transistor logic (3.3 V) pulse that was wired directly to the event input of the APX-15.

The DJI M600P platform was equipped with two different line camera sensors; the Headwall’s visible and near-infrared (VNIR) and shortwave infrared (SWIR) sensors. The VNIR line camera covers 270–273 spectral bands ranging between 398 and 1000 nm with a band width of 2.2 nm. The scan line consists of 640 pixels (samples) and has a detector pitch of 7.5 μm [29]. This sensor has a focal length of 8.2 mm resulting in a field of view of approximately 31° along the scan line. The SWIR line camera covers 166 spectral bands ranging between 900 and 2500 nm with a band width of 9.6 nm. This sensor consists of 384 pixels (samples) along the scan line and has a pixel pitch of 24 μm [30]. The focal length of the SWIR line camera is 24.6 mm resulting in a field of view of 21° along the scan line.

An image of the DJI M600P equipped with the VNIR and SWIR line cameras, along with the orientation of the sensors and IMU body coordinate systems is shown in Fig. 6. The

![Fig. 5. System configuration for the DJI M200 equipped with the RGB and thermal frame cameras.](image-url)
system records a pulse per second (PPS), frame index count per line, and position and orientation along with the Universal Time Coordinated (UTC) time (at 100 Hz), and VNIR/SWIR clock count. The link between the PPS from the APX, frame index count, and UTC time is used to evaluate the timestamp for each scan line. However, a time delay can still exist between the GNSS/INS timestamp and the image exposure time.

Table II outlines the nominal lever arm components and boresight angles for all the systems used in this article. The M200 frame camera system and the M600P line camera system each had one collection date over study sites located at Purdue’s Agronomy Center for Research and Education. Table III shows the flight parameters and dataset description for each imaging system. The M200 UAV system was flown over an agricultural test field with five checkerboard targets and the line camera datasets were acquired over a study site with different geomorphic features, i.e., grass, pavement, and crops, and with 11 checkerboard targets and several highly reflective and hut-shaped targets [see Fig. 7(c) and (d)]. While the checkerboard targets were used as check points for absolute accuracy assessment in the experimental results, remaining ground targets were only deployed for a LiDAR calibration experiment, which is not the focus of this article. The centers of checkerboard targets were surveyed by a Topcon GR-5 GNSS receiver with an accuracy of 2–3 cm. It should be noted that the image coordinates for check points were manually measured. Some sample images/scenes captured by the frame/line cameras are illustrated in Fig. 7. As shown in this figure, all datasets show repetitive image patterns with significant scale differences. These are even more clear in thermal frame and SWIR line datasets, as they are of lower spatial resolutions and also operate in spectral bands that result in low-textured images/scenes. In addition, from Fig. 7, one can note that RGB frame camera has much larger AFOV than the thermal frame sensor. Similarly, the AFOV of VNIR scanner is larger than that for SWIR line camera. Fig. 8 outlines the flight trajectory and configuration of check points for both frame and line camera datasets.

### IV. METHODOLOGY

In this section, the proposed strategies for automated image feature extraction and matching for frame and line cameras onboard UAV remote sensing systems are introduced. The matching strategy used in this article to establish conjugate
features among overlapping frame camera images is based on a traditional SfM technique which is augmented with several layers of matching outlier removal. On the other hand, by taking advantages of partially ortho-rectified images generated for each flight line, a new strategy for automated feature matching for line camera systems is introduced. Last, a general BA procedure for system calibration of frame and line cameras is presented.

A. Automated Feature Extraction and Matching for Frame Camera Systems

In this article, a graphic processing unit (GPU)-based implementation of the SIFT algorithm, known as SiftGPU [31], is used to efficiently extract image features along with their descriptors in a large set of frame camera images. As discussed in Section I, although the SIFT algorithm is designed to be invariant against scaling, in practice this technique results in a high percentage of outliers when dealing with images acquired at significantly different flying heights with repetitive image patterns. In order to increase the reliability of established conjugate SIFT features when applied on images with such challenging conditions, this article implements four layers of matching outlier removal (see Fig. 9):

1) Forward/backward consistency check: In this step, considering two overlapping images, hereafter denoted as left image and right image, first, left-to-right (forward) matching is conducted and an initial set of conjugate features is established. Then, a given pair $(p_1, p_2)$ is considered to be valid if the same pair is established in the right-to-left (backward) matching. This step helps with initial removal of obvious matching outliers.

2) RANSAC-based ROP estimation: Here, the Nister five-point algorithm [32] is augmented with a RANSAC [13] framework to simultaneously estimate ROPs and remove potential matching outliers while using epipolar geometry constraints.

3) Incremental outlier removal: In order to remove more matching outliers arising from the repetitive patterns and scale differences in the imagery, this step of the proposed SfM technique implements an incremental EOP recovery process based on the work presented by He et al. [33]. This EOP recovery step starts with establishing a local coordinate system by selecting a stereo-pair, and then sequentially augmenting the remaining images into the final image block. Image augmentation is achieved using the linear rotation averaging and translation averaging techniques introduced by He et al. [33]. The matching outlier removal step is conducted within the translation averaging process where conjugate points that exhibit large back-projection residuals are detected and removed. Interested readers can refer to [33] for more detailed information about the employed incremental EOP recovery strategy.

4) BA-based outlier removal: In the last step of the SfM framework employed in this article, a BA process is incorporated while using an iterative procedure for detecting and removing more matching outliers. Prior to conducting the BA process, first all remaining SIFT features (i.e., those surviving the prior outlier removal strategies) are tracked among all the involved imagery, and their approximate object coordinates are derived using a multi light-ray intersection procedure. Next, using the SIFT features, their corresponding object coordinates, and camera EOP, a BA process is conducted to minimize the back-projection residuals for all image tie points. At this stage, most SfM pipelines that conduct BA-based outlier removal such as COLMAP [34], simply remove points with large residuals. Although the huge redundancy and geometric strength of the BA process can be assumed to map matching outliers to the corresponding image coordinate residuals, there might be situations where inlier leverage points (i.e., sparse points in key geometric locations) could end up with large residuals. In order to avoid such situations, this article conducts a chi-squared statistical test, where a standardized test of the residuals while considering their variance–covariance matrix is used for removing potential outliers. After removing these outliers, the BA is executed again to refine image EOP as well as reconstructed object points. This procedure is repeated until no points with residuals larger than a predefined threshold remain.

It is worth mentioning that the implemented multiscale image matching strategy for frame camera systems does not exploit any information from the onboard GNSS/INS unit. Thus, this approach is capable of identifying conjugate features even in systems with large deviations of the actual calibration parameters from their nominal values. Fig. 10 shows sample 3-D point clouds generated from—1) Pix4D Mapper Pro and (2) the implemented multiscale approach for the RGB frame dataset acquired at 20 and 40 m flying heights. As illustrated in Fig. 10, although Pix4D Mapper Pro is able to generate more points in the object space, the 3-D reconstruction accuracy is lower when compared to the result from the implemented approach.
More specifically, from the perspective view shown in Fig. 10, one can note that using Pix4D Mapper Pro, images captured at different flying heights result in two separate surfaces. This highlights the challenge in finding conjugate features between images with significant scale differences. One would expect even worse results from existing SfM strategies when dealing with multiscale thermal images due to their more challenging characteristics (e.g., lack of uniquely identifiable features). Therefore, we cannot compare the performance of those approaches with the proposed multiscale matching strategy.

B. Automated Feature Extraction and Matching for Line Camera Systems

As mentioned earlier, in addition to the challenges caused by scale differences, in UAV-based line camera systems, wavy patterns in the generated scenes further complicates the image matching problem. In order to overcome the feature matching problem in raw scenes with wavy patterns, the proposed framework establishes the feature correspondences on partially ortho-rectified scenes of each flight line. An ortho-rectified scene is obtained by correcting the raw scenes for nonsmooth trajectory as well as perspective and camera distortions, if available, such that the generated scene is correctly scaled and corresponds to a map projection throughout the scene. The proposed strategy for automated generation of tie features among line camera scenes is presented in Fig. 11. As shown in this figure, the proposed approach consists of four main steps: partial ortho-rectification of individual flight lines, window-based SIFT matching, SIFT feature tracking, and conversion of ortho-rectified coordinates to raw scene coordinates. Implementation details of these steps are discussed in the remainder of this subsection.

Generation of partially ortho-rectified scenes aims at removing the wavy pattern of the raw scenes, so that corresponding features in different flight lines look similar. The employed ortho-rectification process is depicted in Fig. 12. As illustrated in this figure, in order to appropriately map intensities observed by the camera to their location in the ortho-rectified scene, a rough elevation model (DSM) of the scene in question is required. Since the perspective distortions in line camera scenes are not severe when compared to those for frame camera imagery, possible errors in the used DSM will not significantly affect the quality of the partially ortho-rectified scene.

Given this surface model and using GNSS/INS trajectory information, for each DSM cell, its closest scan line (in 2-D) is determined and denoted as an initial scan line where this cell is believed to be visible. One should note that a DSM cell is visible in its corresponding initial scan line only if the following conditions hold.

1) The scanner is perfectly vertical.
2) There is no variability in the roll, pitch, or yaw in the scanner’s attitude.
In the next step of the proposed framework, the SIFT detector and descriptor algorithm is applied on all partially ortho-rectified scenes. The traditional matching strategy establishes conjugate features by comparing each feature descriptor in one scene with all feature descriptors in the other scene. In some applications such as digital agriculture, due to repetitive image patterns, traditional feature matching algorithms often result in insufficient number of matched features and/or matches with high percentage of outliers [35]. Reducing the search space in such situations could mitigate some of the matching ambiguity problems caused by repetitive patterns/scale difference in imagery. As such, several approaches using data structure techniques and/or trajectory information have been introduced to improve the matching for frame camera imagery [36], [37]. This article exploits the geo-location information of the partially ortho-rectified scenes to reduce the matching search space in line camera scenes. In other words, for each selected SIFT feature in the first scene, its predicted location in the second scene is estimated. Then, among all SIFT features in the second scene, only those which are located inside a search window centered at the predicted point location are considered as candidate conjugate features. The search window size can be determined according to the accuracy of the trajectory information and nominal system calibration parameters, as well as the quality of the utilized rough DSM. Finally, to remove obvious matching outliers, a forward/backward consistency check is conducted. Fig. 14 illustrates the abovementioned process for the two partially ortho-rectified scenes shown in Fig. 13.

Now that conjugate features are established among all stereotypes of ortho-rectified scenes, all SIFT matches are tracked among all involved flight lines. One should note that coordinates of the established conjugate features have been determined with respect to the coordinate systems associated with the ortho-rectified scenes, while for the triangulation process, raw scene coordinates are required. Hence, using the look-up table generated during the ortho-rectification stage, sample and line indices are assigned to each tracked SIFT feature. Fig. 15 shows a sample of tracked SIFT feature on ortho-rectified scenes as well as corresponding raw scenes. Through visual inspection, a good matching quality in both ortho-rectified and raw scenes can be seen. One should note that in contrast to the implemented multiscale matching strategy for frame cameras, the proposed matching algorithm for line camera scenes exploits trajectory information provided by the onboard GNSS/INS unit. Thus, when dealing with line camera systems with significant differences between nominal and actual system calibration parameters, relaxed thresholds, e.g., large search window size, might need to be selected. The robustness of the proposed multiscale matching strategy for line cameras with significant variations of the system calibration parameters from their nominal values will be tested in the results and discussion section of this article.

C. Bundle Block Adjustment

This section presents a general BA procedure with geometric system calibration capabilities for frame and line cameras using automatically derived tie points. In this article, a vector connecting point “$b$” to point “$a$” relative to a coordinate system
Fig. 14. Illustration of the proposed feature matching strategy for a sample stereo-pair of line camera scenes, (a) and (b) detected SIFT features (red dots) on the first and second partially ortho-rectified scenes, respectively, (c) illustration of the matching strategy in zoomed-in regions corresponding to highlighted black boxes in (a) and (b), (d) zoomed-in views of the established conjugate features, and (e) final 558 matched features between the two ortho-rectified scenes.

associated with point “b” is represented as $r^b_i$, and a rotation matrix transforming from coordinate system “a” to coordinate system “b” is represented as $R_{ab}$. The mathematical model for the collinearity principle while considering spatial and rotational system calibration parameters is shown in (1). As mentioned earlier, while both $x, y$ components of image coordinates – $r^{i(t)}_c$ in (1)—in frame cameras have variable values depending on image point location, in line camera systems $x$ is always constant, e.g., $x = 0$ for systems with scan line vertically below the camera perspective center. Despite this difference in image coordinates of line cameras, the overall collinearity equations shown in (1) remain unchanged. An illustration of the collinearity equations is presented in Fig. 16 for frame and line camera systems:

$$r^{m_i}_I = r^{m_i}_{b(t)} + R^b_{b(t)} r^b_c + \lambda (i, c, t) R^m_{b(t)} R^b_c r^{c(t)}_i \tag{1}$$

where

- $r^{m_i}_I$ ground coordinates of the object point $I$,
- $t$ camera exposure time,
vector connecting perspective center to image point $i$ captured by camera $c$ at time $t$, $r^m_{b(t)}$ position of IMU body frame relative to the mapping reference frame at time $t$ derived from the GNSS/INS integration process, $R^m_{b(t)}$ rotation matrix from IMU body frame at time $t$ to the mapping reference frame derived from GNSS/INS integration process, $r^b_c$ lever arm from IMU body frame to camera frame, $R^b_c$ rotation (boresight) matrix from camera frame to IMU body frame, and $\lambda (i, c, t)$ scale factor for point $i$ captured by camera $c$ at time $t$. As mentioned earlier, this article aims at deriving both spatial and temporal calibration parameters for frame and line cameras. LaForest et al. [9] introduced the modified collinearity equations for frame camera systems that incorporate the time delay between the mid-exposure and recorded event marker by the GNSS/INS event marker, whereas time of exposure is recorded for each individual scan line in line camera systems. In the first step of the general BA procedure shown in Fig. 17, the initial exposure time $t_0$ for each tie point is derived from the recorded event marker and timestamp of the corresponding scan line for frame and line cameras, respectively. The initial exposure time $t_0$ is used as the starting time $t'$ in the first iteration. Then, the position and orientation of the IMU body frame at time $t'$ as well as instantaneous linear and angular velocities are computed through an interpolation using the trajectory information (with 200 Hz data rate) provided by GNSS/INS unit. In this process, linear interpolation is used for positional components, whereas rotational parameters are derived through spherical linear interpolation [38]. Next, the BA implements a least squares solution to determine the best estimates of unknown parameters. Once an incremental time delay—$\Delta t'$—is estimated, the initial exposure time is updated to $t_0 + \Delta t'$. In other words, the position and orientation of the IMU body frame as well as the instantaneous linear and angular velocities at time $t_0 + \Delta t'$ are computed for the next iteration of the BA. The iteration procedure is repeated until the incremental time delay is smaller than a predefined threshold. Last, the final time delay is derived through summation of incremental time delays estimated in all iterations, and the
spatial calibration parameters are derived from the last iteration of the BA process.

In order to estimate the system parameters with the least amount of control, previous research has been completed on deriving the optimal/minimal flight configuration which considers any small changes in the system parameters that would impact the target function of the BA [9]. Since changes in lever arm component in the Z direction does not impact the precision of the BA—i.e., wrong estimates of such lever arm component will not impact the BA target function, it cannot be estimated in the absence of vertical control [39]. Therefore, this parameter is not estimated in this article. For this article, the observations include the image coordinate measurements of the check and tie points. The image coordinate measurements for check points were manually measured whereas the image coordinate measurements for the tie points were automatically derived, as explained in the previous section.

V. RESULTS AND DISCUSSION

In this section, the experimental results for the RGB and thermal frame cameras as well as the VNIR and SWIR line cameras are presented. The collective objectives of the experimental results using these four datasets are listed as follows:

1) The first objective is to evaluate the proposed multiscale matching approaches’ ability to establish well-distributed and sufficient number of conjugate features among images/scenes acquired from different flying heights. All four datasets are included in testing this objective.
2) The next objective of the experimental results is to compare the estimated system calibration parameters and absolute accuracy of the reconstruction while using manually measured check points against those using SIFT-based tie points. All four datasets are used to test this objective.
3) The third objective tests the feasibility of the proposed automated calibration strategies when working with systems with significant deviation of actual system parameters from their nominal values while accurately estimating those parameters and ground coordinates of check points. As mentioned earlier, the performance of the proposed strategy for frame cameras does not depend on the quality of the nominal system calibration parameters. Therefore, this objective is tested only for one of the line camera systems, i.e., VNIR dataset.
4) The last objective of the experimental results is to examine the potential use of the proposed multiscale matching strategies in applications other than system calibration, e.g., SfM-based applications, where it is required to reconstruct precise 3-D point clouds from images captured at multiple heights. To test this objective, the relative accuracy of the reconstructed SIFT-based object points is evaluated for the four datasets.

Objective 1—Qualitative and Quantitative Sufficiency of Multiscale Matching Results for Frame and Line Cameras

In this section, multiscale matching results for the four datasets are evaluated in terms of the number of established tie features, distribution of tie features, and the ability of the proposed strategies to identify conjugate features among images acquired from different flying heights. Table IV presents the total number of acquired images for the four datasets, total and average number of established conjugate features in image space, number of reconstructed object points, and average matching repeatability (or average number of images among which a given object point is tracked). As can be seen in this table, the proposed multiscale matching strategy for frame cameras establishes approximately five times more conjugate features for the RGB dataset when compared to the thermal one. This performance disparity is mainly due to the fact that RGB images encompass more distinctive features than the low-textured thermal images [see sample images in Fig. 7(a) and (b)]. Inspecting the matching repeatability criterion reported in Table IV, it can be observed that the average number of images among which an object point is visible is slightly higher for the RGB dataset than the thermal one, i.e., 5.0 versus 4.8. This can be attributed to the higher percentage of overlap and side-lap (due to the larger AFOV) between the RGB images (see Table III).

As explored before with the help of Fig. 7(c) and (d), scenes generated by the SWIR line camera are more challenging for establishing conjugate features when compared with VNIR line camera scenes. However, as reported by the total number of conjugate points in Table IV, almost the same number of SIFT matches are identified for the two datasets. This can be explained by the implemented window-based image matching strategy for line camera scenes, as it can reduce the matching ambiguity which is more significant in the low-textured SWIR camera scenes. Also, matching repeatability for the VNIR dataset is slightly higher than the SWIR dataset due to the increased percentage of side-lap between neighboring VNIR scenes. It is worth mentioning that since two different study sites with different geomorphic features were used for the frame and line camera calibration experiments, no comment can be made on the comparative performance of the proposed multiscale matching strategies for frame and line cameras. However, assuming similar flight configuration over the same study site, one would expect smaller set of conjugate points in line camera datasets due to the absence of overlap among successive scan lines as well as smaller field of view in such cameras.

| Dataset | Total # of images | Total # of conjugate points | # of conjugate points per image | # of object points | Matching repeatability |
|---------|------------------|---------------------------|---------------------------------|-------------------|-----------------------|
| RGB Frame | 450 | 1,250,000 | 2800 | 250,000 | 5.0 |
| Thermal Frame | 450 | 216,000 | 480 | 45,000 | 4.8 |
| VNIR Line | 30,000 | 8,000 | 0.27 | 2,002 | 4.0 |
| SWIR Line | 29,000 | 8,200 | 0.28 | 2,077 | 3.9 |
For evaluating the distribution of the established tie points, Fig. 18 illustrates the image/scene location of all manually measured check points (in red) and automatically derived SIFT points (in blue) in the different images/scenes for (a) RGB frame, (b) thermal frame, (c) VNIR line (left: 20 m and right: 40 m), and (d) SWIR line (left: 20 m and right: 40 m) datasets.

Fig. 19. Percentage of the SIFT-based object points identified in different flying heights for the four datasets.

Specifically, this figure shows the percentage of the SIFT-based object points that are established among images acquired only at a single flying height (20 m in blue and 40 m in orange) as well as this percentage for the object points that are identified and matched among images captured at multiple flying heights (20 m and 40 m in green). Thus, the tie points corresponding to the green bar in Fig. 19 have the major role in avoiding correlations among the system calibration parameters as they tie the two flying heights more efficiently. As illustrated in Fig. 19, in all cases, more than 50% of the SIFT-based object points are derived from both flying heights. Therefore, one can conclude that the proposed multiscale matching strategies for both frame and line camera systems can establish sufficient number of features among images/scenes acquired from different flying heights.

In addition, as images acquired at 40 m flying height have more ground coverage than images captured at 20 m, one might expect those images to include more distinctive features and consequently, result in more SIFT-based conjugate points. However, a closer inspection of the results for the thermal and SWIR cameras in Fig. 19 reveals that for these datasets, the percentage of the features established only in images acquired at 20 m flying height is significantly larger than that at 40 m. This can be attributed to the fact that these cameras, as mentioned earlier, operate in spectral bands that result in images with very low level of details and this condition deteriorates as the spatial resolution of the images decreases. Consequently, in images/scenes captured by thermal/SWIR cameras, it is likely to detect fewer distinctive features in images captured at a higher altitude, i.e., 40 m, and therefore, smaller set of conjugate points can be established among those images.

Objective 2—Reconstruction Accuracy Using the Estimated System Calibration Parameters for Frame and Line Cameras

In this section, using the automatically derived tie points, system calibration parameters, and absolute accuracy of the resulting reconstruction are estimated and compared with those using manual measurements of check points. The nominal values and estimated system calibration parameters using check points and SIFT points for the four datasets are presented in Table V. Standard deviation (STD) of each estimated parameter and the derived square root of a-posteriori variance factor, \( \sigma_0 \), from the BA procedure are also reported. The \( \sigma_0 \) value indicates the quality of fit between observations and estimated unknowns as represented by the mathematical model of the collinearity equations. According to Table V, \( \sigma_0 \) values for all datasets are smaller than 4 pixels, which is an indication of small image residuals as a result of the back-projection process using the estimated system calibration parameters. Although, as expected, the back-projection errors when using accurate manual measurements of check points are smaller than those resulting from automatically derived tie points, the derived \( \sigma_0 \) values using the proposed strategies are still within the expected accuracy of SIFT-based conjugate points (for the BA process in this article, accuracy of manual measurements and SIFT points are considered as 0.5 and 7 pixels, respectively). Looking into the system calibration parameters reported in Table V, one can see that for all datasets, SIFT and check points result in very similar estimates for time delay, lever arm components, and
boresight angles, with slight improvements in STDs of the estimated parameters when using SIFT-based tie points due to the high redundancy that these points produce. Also, from the reported time delays, it can be observed that while there is a significant time delay in frame camera systems (approximately −200 and −270 ms for RGB camera and thermal camera, respectively), time delay estimated for line cameras ranges between +5 and +10 ms, which is considered relatively small. It is worth mentioning that a negative estimated time delay shows that the exposure time is recorded before the actual moment of image capture, whereas a positive time delay is an indication of recording the exposure time after the image is captured.

For an absolute accuracy assessment, the $X/Y/Z$ ground coordinates of the check points are derived by applying the estimated system parameters in a multiray intersection while using the position and orientation of the GNSS/INS unit at the time of exposure. The derived ground coordinates are then compared to the corresponding RTK GNSS measurements. A summary of the mean, STD, and root-mean-squared error (RMSE) of the differences in $X$, $Y$, and $Z$ coordinates of the check points for the four datasets is presented in Table VI. The overall horizontal and vertical components of the RMSE differences while using check points and SIFT points are no more than 1 and 4 cm, respectively, which are considered negligible according to the GSDs of the utilized cameras. Larger vertical component of the RMSE values can be mainly attributed to using inaccurate focal length, resulting in a constant shift as reflected by the large $Z$ mean values for both cases (check and SIFT points). Overall, one can conclude that the proposed multiscale matching strategies for frame and line cameras can lead to system calibration parameters with the same absolute accuracy level as those when using manual measurements of check points.

### Objective 3—Applicability of Proposed Approach for Line Camera System Calibration With Erroneous Nominal Parameters

As presented in Objective 2, depending on the utilized time synchronization technology, a small or large time delay might exist in an imaging system. On the other hand, according to the positional and rotational offset between the onboard camera and GNSS/INS unit, only a small deviation between the actual and nominal mounting parameters is expected (up to a few centimeters/degrees for the lever arm components/boresight angles).

### TABLE V

**SYSTEM CALIBRATION RESULTS USING MANUALLY AND AUTOMATICALLY DERIVED TIE POINTS FOR THE FOUR DATASETS**

| Dataset      | Approach | $\hat\sigma_0$ (pixel) | Time delay $\Delta t$ (ms) | Lever arm $\Delta X$ (cm) | Lever arm $\Delta Y$ (cm) | Bore sight $\Delta \omega$ (°) | Bore sight $\Delta \varphi$ (°) | Bore sight $\Delta \kappa$ (°) |
|--------------|----------|------------------------|-----------------------------|--------------------------|--------------------------|-------------------------------|-------------------------------|-------------------------------|
| RGB Frame    | Check points | 1.55 ± 2.9            | 9.6 ± 1.7                  | 1.2 ± 1.7                | 179.08 ± 0.03           | -0.41 ± 0.04                 | -90.95 ± 0.04                |
|              | SIFT points | 3.48 ± 0.1            | 6.1 ± 0.3                  | 2.5 ± 0.2                | 179.11 ± 0.01           | -0.41 ± 0.01                 | -90.92 ± 0.01                |
| Thermal Frame| Check points | 2.01 ± 7.3            | 5.3 ± 3.1                  | -3.1 ± 3.1               | 178.78 ± 0.06           | -0.24 ± 0.07                 | -90.95 ± 0.11                |
|              | SIFT points | 2.71 ± 6.0            | 5.9 ± 0.3                  | -2.4 ± 0.3               | 178.80 ± 0.01           | -0.20 ± 0.01                 | -90.71 ± 0.01                |
| VNIr Line    | Check points | 0.30 ± 1.5            | -0.01 ± 0.6                | 6.8 ± 0.6                | 0.13 ± 0.01             | 0.36 ± 0.02                  | 179.80 ± 0.03                |
|              | SIFT points | 2.70 ± 0.5            | 0.01 ± 0.3                 | 7.3 ± 0.3                | 0.16 ± 0.01             | 0.35 ± 0.01                  | 179.94 ± 0.02                |
| SWIR Line    | Check points | 0.85 ± 0.8            | -9.5 ± 0.6                 | 4.9 ± 0.6                | 0.25 ± 0.01             | 0.19 ± 0.01                  | 0.04 ± 0.04                  |
|              | SIFT points | 1.86 ± 0.1            | -9.2 ± 0.01                | 4.8 ± 0.01               | 0.23 ± 0.01             | 0.18 ± 0.01                  | 0.16 ± 0.01                  |

### TABLE VI

**ABSOLUTE ACCURACY OF THE SYSTEM CALIBRATION PROCEDURES WHEN USING CHECK POINTS AND SIFT POINTS FOR THE FOUR DATASETS**

| Dataset      | Approach | Mean $X/Y/Z$ (cm) | STD $X/Y/Z$ (cm) | RMSE $X/Y/Z$ (cm) |
|--------------|----------|-------------------|------------------|-------------------|
| RGB Frame    | Check points | 0.3 ± 2.6         | 0.6 ± 1.1        | 4.2 ± 0.7         |
|              | SIFT points | 0.3 ± 2.5         | 0.7 ± 1.0        | 4.2 ± 0.8         |
| Thermal Frame| Check points | 1.0 ± 3.0         | 2.0 ± 1.0        | 8.5 ± 2.0         |
|              | SIFT points | 0.7 ± 2.9         | 1.9 ± 1.0        | 8.2 ± 2.0         |
| VNIr Line    | Check points | 2.5 ± 2.5         | 0.6 ± 1.1        | 3.1 ± 2.5         |
|              | SIFT points | 2.3 ± 2.1         | 0.9 ± 1.0        | 3.6 ± 2.4         |
| SWIR Line    | Check points | 1.6 ± 1.5         | 1.5 ± 1.5        | 8.4 ± 2.1         |
|              | SIFT points | 0.8 ± 0.1         | 2.6 ± 2.1        | 11.8 ± 2.6        |

RMSE Difference (Check points – SIFT points)

| Dataset      | Approach | Mean $X/Y/Z$ (cm) | STD $X/Y/Z$ (cm) | RMSE $X/Y/Z$ (cm) |
|--------------|----------|-------------------|------------------|-------------------|
| RGB Frame    | Check points | 0.3 ± 2.6         | 0.6 ± 1.1        | 4.2 ± 0.7         |
|              | SIFT points | 0.3 ± 2.5         | 0.7 ± 1.0        | 4.2 ± 0.8         |
| Thermal Frame| Check points | 1.0 ± 3.0         | 2.0 ± 1.0        | 8.5 ± 2.0         |
|              | SIFT points | 0.7 ± 2.9         | 1.9 ± 1.0        | 8.2 ± 2.0         |
| VNIr Line    | Check points | 2.5 ± 2.5         | 0.6 ± 1.1        | 3.1 ± 2.5         |
|              | SIFT points | 2.3 ± 2.1         | 0.9 ± 1.0        | 3.6 ± 2.4         |
| SWIR Line    | Check points | 1.6 ± 1.5         | 1.5 ± 1.5        | 8.4 ± 2.1         |
|              | SIFT points | 0.8 ± 0.1         | 2.6 ± 2.1        | 11.8 ± 2.6        |

RMSE Difference (Check points – SIFT points)
from the BA $\hat{\Delta}$, introducing artificial time delay to the $25 \text{ ms}$ interval, for the VNIR dataset captured from $40 \text{ m}$ flying height. From this figure, one can note that as the amount of the introduced artificial time delay increases, more significant distortions are observed in the partially ortho-rectified scenes.

The number of reconstructed object points, $\hat{\sigma}_0$ from the BA procedure, expected time delay, and estimated system calibration parameters along with their STDs for the VNIR dataset with and without artificial time delay are presented in Table VII. For all experiments, a constant search window size ($2 \times 2 \text{ m}$) — which is selected according to the maximum displacement of $0.75 \text{ m}$ resulting from the maximum flying speed ($6 \text{ m/s}$) and maximum artificial time delay introduced ($125 \text{ ms}$) — is used for deriving conjugate features. Looking into the number of reconstructed object points in Table VII, it can be seen that by introducing larger time delays, fewer tie points are identified. This is mainly caused by the distortions in the partially ortho-rectified scenes for the experiments with significant time delay, as shown in Fig. 20. According to the $\hat{\sigma}_0$ values reported in Table VII, although the back-projection errors increase (from $2.7$ to $5.1 \text{ pixels}$) as larger time delays are introduced, these errors are still within the expected accuracy of the SIFT-based tie points. The time delay estimated for the VNIR dataset without an artificial time delay was $10 \text{ ms}$ (shown in Table V and Table VII). Accordingly, by introducing time delays in the range of $25$ to $125 \text{ ms}$ with $25 \text{ ms}$ interval, expected estimated values should range between $35$ to $135 \text{ ms}$ with $25 \text{ ms}$ interval. From the estimated time delays reported in Table VII, one can note that the deviation between the expected and estimated time delays is small (no more than $6 \text{ ms}$) for all cases, except for the experiment with $125 \text{ ms}$ artificial time delay. More specifically, when $125 \text{ ms}$ artificial time delay is introduced to the system, there is a $47 \text{ ms}$ difference between the estimated and expected time delay values. In addition, in this experiment, estimated mounting parameters, particularly boresight angles $\Delta \varphi$ and $\Delta \kappa$, show slightly large deviations from their estimated values when there is no artificial time delay introduced. It is worth mentioning that although the proposed approach does not lead to accurate system parameters when a $125 \text{ ms}$ artificial time delay is introduced, the estimated time delay can be used to generate new ortho-rectified scenes, and consequently, new set of conjugate features with higher percentage of inliers can be derived. This process can be conducted iteratively until accurate system calibration parameters are estimated — i.e., until no significant change in the estimated system parameters is observed between two successive iterations. One should note that the iterative approach described here is different from the currently implemented procedure (discussed in Section IV), in the sense that the current approach involves iterative technique when solving for the time delay without refining the matched features. However, in the proposed strategy for line camera systems with significant time delays, new set of conjugate features is established in each iteration.

Absolute accuracy results for the experiments with different artificial time delays are presented in Table VIII. As reported in this table, except for the experiment with $125 \text{ ms}$ artificial time delay, which shows a vertical absolute accuracy larger than $11 \text{ cm}$, in other cases the overall mean, STD, and RMSE values are very similar to those with significantly smaller time delay in the system (experiment without any artificial time delay).
of the object points is a function of flying height ($\alpha$), between light rays at the location of object points obtained from the BA is reported in Table X. In addition, according to Fig. 21 and Table IX, it can be seen that for all datasets, although the maximum $\alpha$ angles for check points and SIFT points are very similar, the mean $\alpha$ values ($\bar{\alpha}$) for the former is larger. This shows that although the proposed strategies are able to establish objects points whose intersection quality is as high as intersection quality of check points, the majority of SIFT-based tie points are established only among images with high percentage of overlap, i.e., short baselines. This is an indication of a relatively low matching repeatability for SIFT-based features (see Table IV), particularly in frame camera datasets. For example, while each ground target in the thermal dataset was manually measured in an average of 39 images, each SIFT-based object point was established only in five images. These values were nine images for check points and four images for SIFT points in the SWIR dataset, indicating that using the GNSS/INS information, the proposed multiscale matching strategy for line cameras showed better performance in terms of matching repeatability when compared to the implemented matching approach for frame cameras. Also from Table IX, one can note that in case of check points, the mean $\alpha$ ($\bar{\alpha}$) values for frame camera datasets are larger than those for line cameras. This is mainly caused by the fact that there is no overlap between line camera scenes in the flight direction, and thus, only light rays from neighboring flight lines can contribute to the 3-D reconstruction process when using such cameras.

Having discussed the quality of intersection for the four datasets, the relative accuracy—precision—of the object points is presented below. The relative accuracy is evaluated by examining the STDs of the check/SIFT point coordinates derived from the BA process. One should note that the planimetric precision of the object points is a function of flying height ($H$) and the precision of the image coordinate measurement ($\sigma_{xy}$). On the other hand, vertical precision can be determined from $H$, $\sigma_{xy}$, and the baseline between images. A summary of mean STD—relative accuracy measure—of the check points and SIFT-based object points obtained from the BA is reported in Table X. In addition, to better understand the impact of intersection quality ($\alpha$ angle) on the precision of the object space, relative accuracy of the SIFT points with $\alpha$ angle larger than their corresponding mean value ($\bar{\alpha}$), is also reported in Table X. Looking into the $XYZ$ mean STDs in this table, one can note that for all datasets, check points show a better relative accuracy than SIFT points, especially in

Furthermore, through adopting the abovementioned iterative procedure for line camera systems with significant time delays, we would expect to achieve high absolute accuracy in the experiment with 125 ms artificial time delay. It is worth mentioning that according to the current time synchronization technology used in line camera imaging systems, time delay with such magnitude is rare and atypical in practice. Overall, one can conclude that the proposed approach accurately estimates the system calibration parameters and achieves similar absolute accuracy whether the time delay in the system is minimal or significant.

Objective 4—Feasibility of Using the Proposed Multiscale Matching Strategies for 3-D Reconstruction

The last objective of the experimental results is to examine the potential use of the proposed multiscale matching strategies for generating 3-D point clouds with high precision from UAV-based images acquired at multiple flying heights. This test is conducted by looking into the relative accuracy of the object space derived from the BA process. However, before evaluating the relative accuracy, quality of intersection for the reconstructed object points is examined by measuring the maximum intersection angle, $\alpha$, between light rays at the location of object points. In general, the larger the magnitude of the base-to-height ratio, the larger $\alpha$ and the better the intersection geometry will be. Fig. 21 illustrates histograms of $\alpha$ angles for SIFT-based object points for the four datasets, while representing the mean $\alpha$ angles for SIFT and check points using green and blue lines, respectively. Also, minimum, maximum, mean, and STD values of $\alpha$ angles for the four datasets are reported in Table IX.
Fig. 21. Histogram of $\alpha$ angle for SIFT-based object points for (a) RGB frame, (b) thermal frame, (c) VNIR line, and (d) SWIR line camera datasets. Green and blue lines represent the mean $\alpha$ angles for SIFT and check points, respectively.

**TABLE IX**

| Dataset   | Approach       | Minimum $\alpha$ (°) | Maximum $\alpha$ (°) | Mean $\alpha$ (°) | STD $\alpha$ (°) |
|------------|----------------|----------------------|----------------------|-------------------|------------------|
| RGB Frame  | Check points   | 43.6                 | 46.8                 | 45.4              | 1.1              |
|            | SIFT points    | 1.4                  | 39.7                 | 14.0              | 5.7              |
| Thermal Frame | Check points | 28.7                 | 33.1                 | 30.8              | 1.6              |
|            | SIFT points    | 0.2                  | 28.6                 | 13.0              | 5.1              |
| VNIR Line  | Check points   | 15.7                 | 25.4                 | 21.4              | 3.7              |
|            | SIFT points    | 2.0                  | 25.6                 | 13.5              | 3.4              |
| SWIR Line  | Check points   | 10.1                 | 18.6                 | 14.5              | 2.7              |
|            | SIFT points    | 3.2                  | 18.1                 | 10.2              | 2.7              |

**TABLE X**

| Dataset   | Approach       | Mean Standard Deviations |
|------------|----------------|--------------------------|
|            |                | $X$ (cm)  | $Y$ (cm)  | $Z$ (cm)  |
| RGB Frame  | Check points   | 1.1       | 1.2       | 4.2       |
|            | SIFT points    | 4.9       | 5.6       | 19.9      |
|            | SIFT points with $\alpha > \bar{\alpha}$ | 2.6 | 3.0 | 10.3 |
| Thermal Frame | Check points | 2.4       | 2.4       | 12.7      |
|            | SIFT Points    | 5.2       | 6.4       | 45.9      |
|            | SIFT points with $\alpha > \bar{\alpha}$ | 3.2 | 3.8 | 23.8 |
| VNIR Line  | Check points   | 0.8       | 0.7       | 5.2       |
|            | SIFT points    | 1.5       | 1.1       | 15.4      |
|            | SIFT points with $\alpha > \bar{\alpha}$ | 1.2 | 1.0 | 12.1 |
| SWIR Line  | Check points   | 0.9       | 0.7       | 7.1       |
|            | SIFT points    | 4.2       | 2.5       | 24.8      |
|            | SIFT points with $\alpha > \bar{\alpha}$ | 3.3 | 2.2 | 18.9 |
vertical direction. While the better planimetric precision can be attributed to the higher accuracy of manual measurements compared to automated extraction of SIFT features, lower vertical relative accuracy for SIFT points is mainly caused by the weaker intersection geometry, i.e., small $\alpha$ angles. Examining the $X/Y$ mean STDs reported in Table X, it is worth mentioning that although the planimetric precision of SIFT points are not as high as that for the check points, they are no more than four times that of the GSD of the original images at the 40 m flying height. Furthermore, from Table X, it can be seen that by removing SIFT points with weaker intersection geometry, planimetric and vertical relative accuracy can be improved for all datasets. These improvements, which are more pronounced in the vertical direction, show the impact of intersection geometry on the precision of the reconstructed object points. Based on the relative accuracy evaluation in this section, one can conclude that the proposed multiscale matching strategies have the potential to be used in SFM-based applications and produce point clouds with high planimetric relative accuracy. Regarding the vertical precision of the reconstructed object space, further improvements can be investigated, especially for the proposed approach for frame cameras by using the available GNSS/INS information, to increase the matching repeatability and consequently improve the intersection quality.

VI. CONCLUSION

This research presented a methodology for reliable multiscale automated feature extraction and matching for both frame and line camera systems. The automatically derived tie points were then used in a general GNSS/INS-assisted BA with system self-calibration to produce accurate system parameters. The key motivation for such development is mitigating the limitations of existing UAV system calibration techniques for automated estimation of spatial and temporal system calibration parameters for both frame and line cameras. Results showed that the matching approach paired with the automated system calibration presented in this research was capable of producing the same level of absolute accuracy compared to using manually measured tie points for both frame and line camera systems. Additionally, the results indicated that the automated system calibration approach presented in this article can be applied to systems with significant and minimal time delay discrepancies and still achieve reliable estimates of system parameters and absolute accuracy. In addition, the relative accuracy analysis showed that the proposed multiscale matching algorithms can also be used in SFM-based applications and produce 3-D object space with high/moderate planimetric/vertical precision. In conclusion, this article presented an automated spatial and temporal calibration approach that produces accurate system calibration parameters without the need for GCPs or manually measured tie points for both frame and line camera systems. Future work will focus on simultaneously integrating frame and line camera images as well as LiDAR data in the feature matching and BA procedure. Also, other algorithms such as SURF will be explored as feature matching alternatives.

ACKNOWLEDGMENT

The authors would like to thank Ali Masjedi, Taojun Wang, Karoll Quijano, and John Evan Flatt for their work on system integration, data collection, and data processing; Radhika Ravi for her helpful comments during manuscript preparation; and Dr. Crawford for her valuable input throughout this work. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

REFERENCES

[1] A. Habib, Y. Han, W. Xiong, F. He, Z. Zhang, and M. Crawford, “Automated ortho-rectification of UAV-based hyperspectral data over an agricultural field using frame RGB imagery,” Remote Sens., vol. 8, 2016, Art. no. 796.
[2] M. L. Buchaillot et al., “Evaluating maize genotype performance under low nitrogen conditions using RGB UAV phenotyping techniques,” Sensors, vol. 19, 2019, Art. no. 1815.
[3] Y. Chen, J. Ribera, C. Boomsma, and E. Delp, “Locating crop plant centers from UAV-based RGB imagery,” in Proc. IEEE Int. Conf. Comput. Vis. Workshops, 2017, pp. 2030–2037.
[4] L. Zhang, A. Masjedi, J. Zhao, and M. M. Crawford, “Prediction of sorghum biomass based on image based features derived from time series of UAV images,” IEEE Int. Geosci. Remote Sens. Symp., pp. 6154–6157, 2017.
[5] A. Masjedi et al., “Sorghum biomass prediction using UAV-based remote sensing data and crop model simulation,” in IEEE Int. Geosci. Remote Sens. Symp., 2018, pp. 7719–7722.
[6] A. Mateas et al., “Intercomparison of UAV, aircraft and satellite remote sensing platforms for precision viticulture,” Remote Sens., vol. 7, pp. 2971–2990, 2015.
[7] M. Blázquez, “A new approach to spatio-temporal calibration of multisensor systems,” in Proc. Int. Archives Photogrammetry, Remote Sens. Spatial Inf. Sci., Beijing, China, 2008, pp. 3–13.
[8] M. Rehak and J. Skaloud, “Time synchronization of consumer cameras on micro aerial vehicles,” ISPRS J. Photogramm. Remote Sens., vol. 123, pp. 114–123, 2017.
[9] H. LaForest, F. Zhou, and A. Habib, “New strategies for time delay estimation during system calibration for UAV-based GNSS/INS-assisted imaging systems,” Remote Sens., vol. 11, no. 15, Aug. 2019, Art. no. 1811.
[10] A. Habib, T. Zhou, A. Masjedi, Z. Zhang, J. E. Platt, and M. Crawford, “Bore sight calibration of GNSS/INS-assisted push-broom hyperspectral scanners on UAV platforms,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 11, no. 5, pp. 1734–1749, May 2018.
[11] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” Int. J. Comput. Vis., vol. 60, no. 2, pp. 91–110, 2004.
[12] H. Bay, T. Tuytelaars, and L. van Gool, “Surf: Speeded up robust features,” in European Conference on Computer Vision, Berlin, Heidelberg: Springer, 2006, pp. 404–417.
[13] M. A. Fischler and R. C. Bolles, “Graphics and image processing random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography,” Commun. Assoc. Comput. Mach., vol. 24, no. 6, pp. 381–395, 1981.
[14] A. F. Habib, M. Morgan, and Y. R. Lee, “Bundle adjustment with self-calibration using straight lines,” Photogramm. Record., vol. 17, pp. 635–650, 2002.
[15] P. Rathnayaka, S. H. Back, and S. Y. Park, “Calibration of a different field-of-view stereo camera system using an embedded checkerboard pattern,” in Proc. Int. Conf. Comput. Vis. Theory Appl., vol. 5, 2017, pp. 294–300.
[16] Z. Zhang, “A flexible new technique for camera calibration,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 1, pp. 1330–1334, Nov. 2000.
[17] Q. Pentek, P. Kennel, T. Allouis, C. Fiorio, and O. Strauss, “A flexible targetless lidar–GNSS/INS–camera calibration method for UAV platforms,” ISPRS J. Photogramm. Remote Sens., vol. 166, pp. 294–307, Aug. 2020.
[18] R. Ravi, J. Y. Lin, M. Elbalhmasawy, T. Shamseldin, and A. Habib, “Simultaneous system calibration of a multi-LiDAR multicamera mobile mapping platform,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 11, no. 5, pp. 1694–1714, May 2018.
[19] K. W. Chiang, M. L. Tsai, E. S. Naser, A. Habib, and C. H. Chu, “A new calibration method using low cost MEM IMUs to verify the performance of UAV-borne MMS payloads,” Sensors, vol. 15, pp. 6560–6585, 2015.
Seyed Meghdad Hasheminasab received the B.S. and M.S. degrees in geomatics engineering from the K.N. Toosi University of Technology, Tehran, Iran, in 2013 and 2016, respectively. He is currently working toward the Ph.D. degree in civil engineering (geomatics) at Purdue University, West Lafayette, IN, USA. His research interests include image/LiDAR-based 3-D reconstruction using mobile mapping systems, structure from motion, and remote sensing image matching and registration.

Tian Zhou received the B.S. degree in remote sensing from Wuhan University, Wuhan, China, in 2015. He is currently working toward the Ph.D. degree in geomatics at Lyles School of Civil Engineering, Purdue University, West Lafayette, IN, USA. His research interests include sensor calibration, 3-D reconstruction from UAV-based imagery, hyperspectral ortho-rectification, and LiDAR point cloud processing.

Lisa M. LaForest received the B.S. degree in mathematics from the University of Missouri, Columbia, MO, USA, in 2013, and the M.S. and Ph.D. degrees in geomatics engineering from Purdue University, West Lafayette, IN, USA, in 2016 and 2020. She has been working with the National Geospatial Intelligence Agency, since 2011. Her research interest includes satellite and UAV sensor modeling with a focus in improving absolute accuracy.

Ayman Habib received the M.Sc. and Ph.D. degrees in photogrammetry from The Ohio State University, Columbus, OH, USA, in 1993 and 1994, respectively. He is currently a Professor in geomatics with Lyles School of Civil Engineering, Purdue University, West Lafayette, IN, USA. His research interests include the fields of terrestrial and aerial mobile mapping systems, modeling the perspective geometry of non-traditional imaging scanners automatic matching and change detection, utilizing low-cost imaging systems for infrastructure monitoring and biomedical applications, incorporating analytical and free-form linear features in various photogrammetric orientation procedures, object recognition in imagery, UAV-based 3-D mapping, and integrating photogrammetric data with other sensors/datasets (e.g., GNSS/INS, GIS databases, multispectral sensors, and LiDAR).

[20] M. Rehak and J. Skaloud, “Fixed-wing micro aerial vehicle for accurate corridor mapping,” ISPRS Ann. Photogrammetry, Remote Sens. Spatial Inf. Sci., vol. 2, 2015.

[21] S. Kocaman, L. Zhang, A. Gruen, and S. Corporation, “Self-calibrating triangulation of airborne linear array CCD cameras,” in Proc. Int. Calibration Orientation Workshop, 2006.

[22] A. F. Habib, K. I. Bang, S.-W. Shin, and E. Mitishita, “Lidar system self-calibration using planar patches from photogrammetric data,” in Proc. 5th Int. Symp. Mobile Mapping Technol., May 2007.

[23] A. S. Gneeniss, J. P. Mills, and P. E. Miller, “In-flight photogrammetric camera calibration and validation via complementary lidar,” ISPRS J. Photogramm. Remote Sens., vol. 100, pp. 3–13, 2015.

[24] T. Zhou, S. M. Hasheminasab, R. Ravi, and A. Habib, “LiDAR-aided interior orientation parameters refinement strategy for consumer-grade cameras onboard UAV remote sensing systems,” Remote Sens., vol. 12, 2020, Art. no. 2268.

[25] “Matrice 200 User Manual,” 2018. [Online]. Available: https://dl.dji.com/downloads/M200/20180910/M200_User_Manual_EN.pdf

[26] “Matrice 600 Pro User Manual,” 2017. [Online]. Available: https://dl.dji.com/downloads/m600%20pro/20180417/Matrice_600_Pro_User-Manual_v1.0_EN.pdf

[27] “APX. Trimble APX-15UAV(V2)—Datasheet,” 2017. [Online]. Available: https://www.applanix.com/downloads/products/specs/APX15_DS_NEW_0408_YW.pdf

[28] “FLIR Duo Pro R User Guide,” 2018. [Online]. Available: https://www.flir.com/globalassets/importedassets/document/duo-pro-r-user-guide-v1.0.pdf

[29] “Nano-Hyperspec Imaging Sensor — Datasheet,” 2018. [Online]. Available: http://www.analytiq.co.uk/wp-content/uploads/2016/03/nano-hyperspec-datasheet.pdf

[30] “Micro-Hyperspec imaging sensors — Datasheet,” 2018. [Online]. Available: https://cdn2.hubspot.net/hubfs/145999/June%202018%20Collateral/MicroHyperspec0418.pdf

[31] C. Wu, “A GPU implementation of scale invariant feature transform (SIFT),” Accessed: Aug. 28, 2020, 2007, [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.214.6333

[32] D. Nistér, “An efficient solution to the five-point relative pose problem,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 26, no. 6, pp. 756–770, Jun. 2004.

[33] J. L. Schonberger and J. M. Frahm, “Structure-from-motion revisited,” in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2016, pp. 4104–4113.

[34] S. M. Hasheminasab, T. Zhou, and A. Habib, “GNSS/INS-assisted structure from motion strategies for UAV-based imagery over mechanized agricultural fields,” Remote Sens., vol. 12, no. 3, Feb. 2020, Art. no. 351.

[35] Y. Sun, L. Zhao, S. Huang, L. Yan, and G. Dissanayake, “L2-SIFT: SIFT feature extraction and matching for large images in large-scale aerial photogrammetry,” ISPRS J. Photogramm. Remote Sens., vol. 91, pp. 1–16, 2014.

[36] S. Jiang and W. Jiang, “On-board GNSS/IMU assisted feature extraction and matching for oblique UAV images,” Remote Sens., vol. 9, no. 8, 2017, Art. no. 813.

[37] K. Shoemaker, “Animating rotation with quaternion curves,” in Proc. 12th Annus. Conf. Comput. Graph. Interact. Techn., 1985, pp. 245–254.

[38] R. Ravi, Y. J. Lin, M. Elbahnasawy, T. Shamseldin, and A. Habib, “Bias impact analysis and calibration of terrestrial mobile LiDAR system with several spinning multibeam laser scanners,” IEEE Trans. Geosci. Remote Sens., vol. 56, no. 9, pp. 5261–5257, Sep. 2018.