Tightly-Coupled Camera/LiDAR Integration for Point Cloud Generation from GNSS/INS-assisted UAV Mapping Systems

Tian Zhou [1], Seyyed Meghdad Hasheminasab [1], Ayman Habib [1,*]

[1] Lyles School of Civil Engineering, Purdue University, West Lafayette, IN 47907, USA; zhou732@purdue.edu (T.Z.); hashemin@purdue.edu (S.M.H.); ahabib@purdue.edu (A.H.)

* Correspondence: ahabib@purdue.edu

Abstract

Unmanned aerial vehicles (UAVs) equipped with integrated global navigation satellite systems/inertial navigation systems (GNSS/INS) together with cameras and/or LiDAR sensors are being widely used for topographic mapping in a variety of applications such as precision agriculture, coastal monitoring, and archaeological documentation. Integration of image-based and LiDAR point clouds can provide a comprehensive 3D model of the area of interest. For such integration, ensuring a good alignment between data from the different sources is critical. Although many works have been conducted on this topic, there is still a need for a rigorous integration approach that minimizes the discrepancy between camera and LiDAR data caused by inaccurate system calibration parameters and/or trajectory artifacts. This study proposes an automated tightly-coupled camera/LiDAR integration workflow for GNSS/INS-assisted UAV systems. The proposed strategy is conducted in three main steps. First, an image-based point cloud is generated using a LiDAR/GNSS/INS-assisted structure from motion (SfM) strategy. Then, feature correspondences between image-based and LiDAR point clouds are

automatically identified. Finally, an integrated-bundle adjustment procedure including image points, LiDAR raw measurements, and GNSS/INS information is conducted to minimize the discrepancy between point clouds from different sensors while estimating system calibration parameters and refining the trajectory information. The proposed SfM strategy and integration framework are evaluated using five datasets. The SfM results show that using LiDAR data can facilitate feature matching and further increase the number of reconstructed 3D points. The experimental results also illustrate that the developed automated camera/LiDAR integration strategy is capable of accurately estimating system calibration parameters to achieve good alignment among camera/LiDAR data from single/multiple systems. An absolute accuracy in the range of 3 to 5 cm is achieved for the image/LiDAR point clouds after the integration process.

Keywords: Camera/LiDAR integration, bundle adjustment, unmanned aerial vehicles, structure from motion, GNSS/INS-assisted mapping, system calibration.

1. Introduction

Topographic mapping using unmanned aerial vehicles (UAVs) is of increasing interest for a variety of applications due to their capability to carry advanced sensors and collect high temporal and spatial resolution data (Nex and Remondino, 2014). RGB cameras and LiDAR scanners are the two most commonly used sensors for UAV-based 3D mapping. Sparse, image-based 3D reconstruction is often conducted via structure from motion (SfM) techniques and can provide spectral information of the area of

interest (Westoby et al., 2012; Fonstad et al., 2013; Schonberger and Frahm, 2016). Assuming that an integrated global navigation satellite system and inertial navigation system (GNSS/INS) unit is available onboard, LiDAR sensors can directly provide dense 3D point clouds – but without spectral/color information – of the area in question. Due to their complementary characteristics, integration of camera and LiDAR sensors can overcome their individual limitations and result in well-aligned products from the two sources including orthophoto, 3D point cloud, digital surface model (DSM)/digital elevation model (DEM), etc.

Aligning data from different sensors requires sensor frames to be geo-referenced in the same coordinate system. In order to derive accurate 3D geospatial information, it is necessary to establish the interior and exterior orientation of the utilized sensors. While the interior orientation parameters (IOP) of LiDAR sensors – often provided by the manufacturer – are relatively accurate and stable, consumer-grade cameras require frequent camera calibration due to the instability of their IOP (including principal distance, principal point coordinates, and lens distortions) (Hastedt and Luhmann, 2015). For camera and LiDAR units mounted on GNSS/INS-assisted UAV systems, exterior orientation parameters (EOP), which define the position and orientation of the sensor at the moment of exposure/firing in a mapping frame, can be estimated using trajectory information. To do so, mounting parameters, which include lever arm components and boresight angles, relating the Inertial Measurement unit (IMU) body

63 frame of the GNSS/INS to the sensor coordinate frame need to be estimated. It is
64 evident that for deriving products with high geometric accuracy, accurate system
65 calibration parameters (i.e., IOP of the sensors and mounting parameters) and trajectory
66 are key prerequisites.

67 Extensive efforts have been conducted on estimating the system calibration of
68 camera and LiDAR units mounted on remote sensing platforms. Most of these system
69 calibration strategies are conducted in indoor environments with the help of special
70 ground targets (Guindel et al., 2017; Pusztai et al., 2018; Xie et al., 2018; Zhou et al., 2018;
71 Cai et al., 2020). In these approaches, features corresponding to the targets are first
72 extracted from both sensors. By minimizing the discrepancy between conjugate
73 features, system calibration parameters can be estimated through a non-linear least-
74 squares adjustment (LSA) procedure. Other than the system calibration-based
75 approaches, data-driven strategies – i.e., point cloud registration – are widely used for
76 integrating 3D point clouds generated by camera and LiDAR systems. The most
77 popular approach for point cloud registration is the iterative closest point (ICP)
78 algorithm (Besl and McKay, 1992; Chen and Medioni, 1992). In this approach,
79 correspondences between the point clouds are established iteratively while
80 continuously refining the transformation parameters between the reference frames of
81 the involved point clouds. However, only the relative accuracy between two point
82 clouds can be guaranteed using registration strategies. Moreover, error sources (e.g.,

point cloud deformation caused by inaccurate camera IOP and/or trajectory) cannot be dealt with when using a rigid spatial transformation model.

Over the past few years, camera/LiDAR integration onboard mobile mapping systems (MMS) equipped with navigation sensors is attracting more attention due to recent advancement in GNSS/INS technology. Some researchers have investigated camera/LiDAR integration within the system calibration. Assuming that a well-calibrated LiDAR system is available, Gneeniss et al. (2015) and Zhou et al. (2020) utilized control information derived from LiDAR data for camera system calibration parameters estimation through a GNSS/INS-assisted BA to improve the alignment of 3D point clouds derived from the imaging and ranging sensors. Ravi et al. (2018a) proposed a target-based system calibration method for a GNSS/INS-assisted multi-LiDAR multi-camera MMS platform. Corresponding features of deployed targets were manually measured from LiDAR point clouds and camera images. The mounting parameters related to all sensors were then estimated through a non-linear LSA by minimizing the distance between 3D coordinates among conjugate features extracted from LiDAR and image data. Although the proposed calibration strategy was able to produce highly accurate mounting parameters for both LiDAR units and cameras, it requires deployment of targets in the calibration site as well as manual measurements of those targets. This process is computationally expensive and labor intensive. Also, the approach assumes that the GNSS/INS trajectory is of high quality.

Several fully-automated approaches were developed for multi-modal system calibration. In these methods, conjugate features in image and LiDAR data were established automatically. Li et al. (2019) proposed a two-step strategy for the integration of image-based and LiDAR point clouds acquired from UAVs. In the first step, camera system calibration parameters and trajectory were refined. More specifically, SfM-derived image-based point cloud, camera IOP and mounting parameters, as well as raw inertial measurements and GNSS observations were included in a single BA procedure to derive accurate camera calibration parameters and a refined high-frequency trajectory. In the second stage, an image-based point cloud generated from the first step was used as reference. LiDAR mounting parameters were then estimated by iteratively minimizing the discrepancy between depth maps derived from the image-based point cloud and LiDAR data. Similarly, Pentek et al. (2020) conducted camera/LiDAR integration in two steps. LiDAR system calibration was first conducted using a LiDAR strip adjustment strategy (Glira et al., 2016). LiDAR mounting parameters were estimated by minimizing the discrepancy between LiDAR point clouds from different flight lines, referred to as LiDAR strips. Then, the LiDAR point cloud generated using refined system calibration parameters was adopted as control information for camera system calibration. A simple metric was proposed in this approach to evaluate the misalignment among conjugate image points caused by inaccurate camera mounting parameters. For a pair of conjugate image points derived

123 through an image matching stage, the corresponding 3D coordinates for each image
124 point were estimated through intersection between the image light ray and LiDAR
125 point cloud. By minimizing the distance between conjugate 3D points, camera
126 mounting parameters were estimated. Although the efficiency of camera mounting
127 parameters estimation was improved, this method did not consider the fact that there
128 might be an error in the available trajectory. Moreover, both integration strategies
129 proposed by Li et al. (2019) and Pentek et al. (2020) calibrate LiDAR and camera sensors
130 sequentially – i.e., the products from the first calibrated sensor are used as reference to
131 calibrate another one. Thus, residual errors in the first calibration would lead to more
132 errors in the second calibration.

133 Glira et al. (2019) proposed a tightly-coupled camera and LiDAR integration process
134 that simultaneously estimates all involved system calibration parameters and refines
135 trajectory for UAV systems. In this process, SfM was first conducted to derive conjugate
136 image points and image-based point cloud. LiDAR point clouds from all flight lines
137 were reconstructed using available trajectory and system calibration parameters. 3D
138 point to 3D point correspondences among all point clouds were automatically identified
139 by finding the closest points. Lastly, an integrated-BA that includes LiDAR raw
140 measurements, conjugate image points, and trajectory was conducted to estimate the
141 system calibration parameters and refine the trajectory. The point cloud generation,

correspondence establishment, and BA steps were conducted iteratively until changes in the sum of squared residuals within the BA is smaller than a predefined threshold.

The strategy proposed by Glira et al. (2019) is a fully automated, tightly-coupled camera/LiDAR integration process where trajectory can be refined as well. However, the process of finding 3D point to 3D point correspondences is computationally expensive. More specifically, the normal vector of each 3D point is evaluated using its neighboring points. Therefore, down-sampling the LiDAR point cloud is essential. Also, minimizing the distance between 3D points from different LiDAR strips is not rigorous especially when the noise level in the LiDAR data is high. This is illustrated in Figure 1, where two sets of LiDAR point clouds from different flight lines with a constant Z discrepancy between them are used as an example. As shown in Figure 1, based on the 3D point to 3D point matching strategy, for each LiDAR point from flight line a, the closest point from flight line b is identified as a correspondence. This leads to a situation that only the upper layer (circled in blue in Figure 1a) of the LiDAR point cloud from flight line b is selected. By minimizing the discrepancy of these established correspondences, the point cloud from flight line a will be aligned with the upper layer of the point cloud from flight line b, as illustrated in Figure 1b. This misalignment will become more severe as the noise level in LiDAR point cloud increases. This problem can be resolved by minimizing the discrepancy between conjugate planar patches. In this case, the noise level in LiDAR point clouds will not affect the performance of the

integration process. It can be seen from Figure 1c that the two LiDAR point clouds are fitted to a common plane as presented by the black dash line. In addition to the limitation in establishing feature correspondences, when incorporating trajectory information in the BA, Glira et al. (2019) modeled errors in trajectory by a cubic spline function. This will introduce many additional unknowns and equations into the BA.

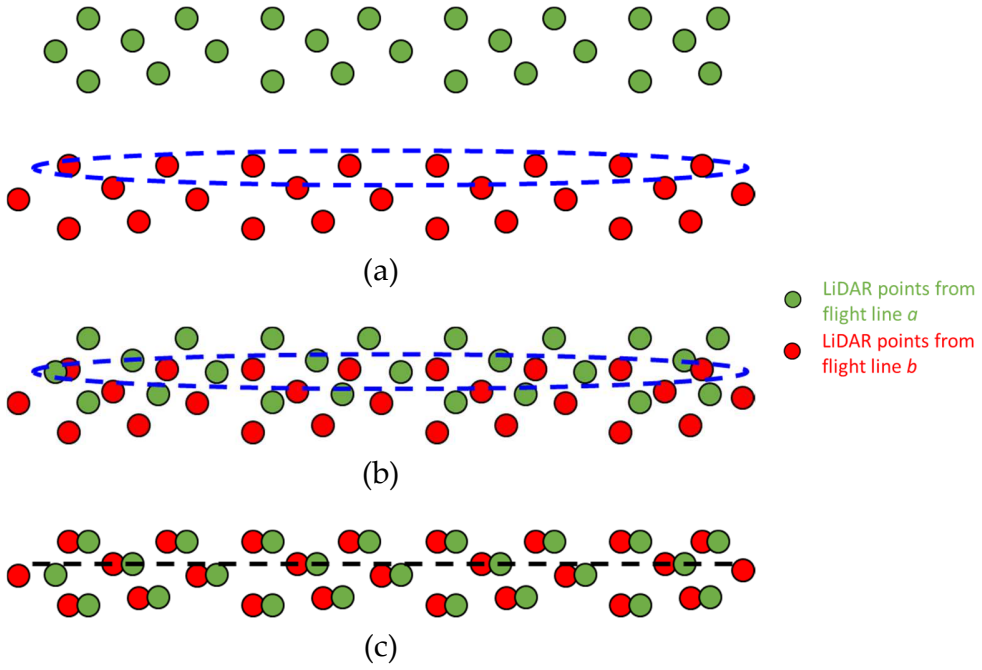


Figure 1. An example of aligning (a) two sets of LiDAR point clouds with Z discrepancy (in side-view) by minimizing (b) 3D point to 3D point distance or (c) planar patch to planar patch discrepancy.

To overcome the above-mentioned limitations in the state-of-the-art techniques, this study provides an automated tightly-coupled camera/LiDAR integration workflow for GNSS/INS-assisted UAV systems. In this paper, the term "tightly-coupled" refers to the fact that the integration process is conducted at the raw data level instead of product level (i.e., loosely-coupled). In the proposed approach, image-based/LiDAR point

clouds are first generated. In this process, LiDAR data is used in a GNSS/INS-assisted SfM procedure to assist image-based 3D reconstruction. Then, the following two steps are adopted iteratively until predefined convergence criteria are met: i) correspondences between image-based and LiDAR point clouds are identified, and ii) an integrated-bundle adjustment (BA) procedure is conducted to estimate the involved system calibration parameters for the camera and LiDAR sensors while refining the trajectory. Once converged, products from both sensors (3D point clouds in this study) are generated using the final system calibration parameters and refined trajectory. The proposed approach has several advantages: i) the workflow does not require deployment of special targets or manual intervention, ii) the involved computation is efficient, iii) the process supports the introduction of GCPs as well as control point clouds, and iv) the integration of multiple platforms is straightforward. The main contributions of this work include:

- Proposing a LiDAR/GNSS/INS-assisted SfM strategy to improve the performance and efficiency of traditional SfM pipelines;
- Presenting an automated workflow, including an integrated-BA model, that utilizes imagery, raw LiDAR measurements, and trajectory information to generate well-aligned image-based/LiDAR point clouds through system calibration and trajectory refinement;

- Presenting a camera/LiDAR integration process that is applicable for aligning multiple sensors/systems and incorporating ground control points (GCPs) and control point clouds to ensure the absolute accuracy of derived products.

The remainder of the paper is structured as follows: Section 2 introduces the proposed automated camera/LiDAR integration workflow. Section 3 describes the adopted UAV systems and datasets used in this study. Section 4 presents the experimental results and discussion. Finally, Section 5 provides the conclusions and recommendations for future work.

2. Camera/LiDAR Data Integration Methodology

In this section, the proposed workflow for tightly-coupled camera/LiDAR integration is introduced. Figure 2 illustrates the workflow of the proposed integration strategy. As shown in this figure, the proposed framework starts with LiDAR and image-based point cloud generation. Then, in the iterative integration stage, automated identification of camera/LiDAR correspondences and an integrated-bundle adjustment (BA) procedure are conducted. The former utilizes image-based object points as seed points for identifying corresponding planar patches in the LiDAR point clouds. To ensure the alignment among the image-based and LiDAR point clouds, an integrated-BA procedure with system calibration and trajectory refinement capabilities for imaging and ranging sensors is adopted. The iterative process will terminate when changes in system calibration parameters are smaller than a predefined threshold.

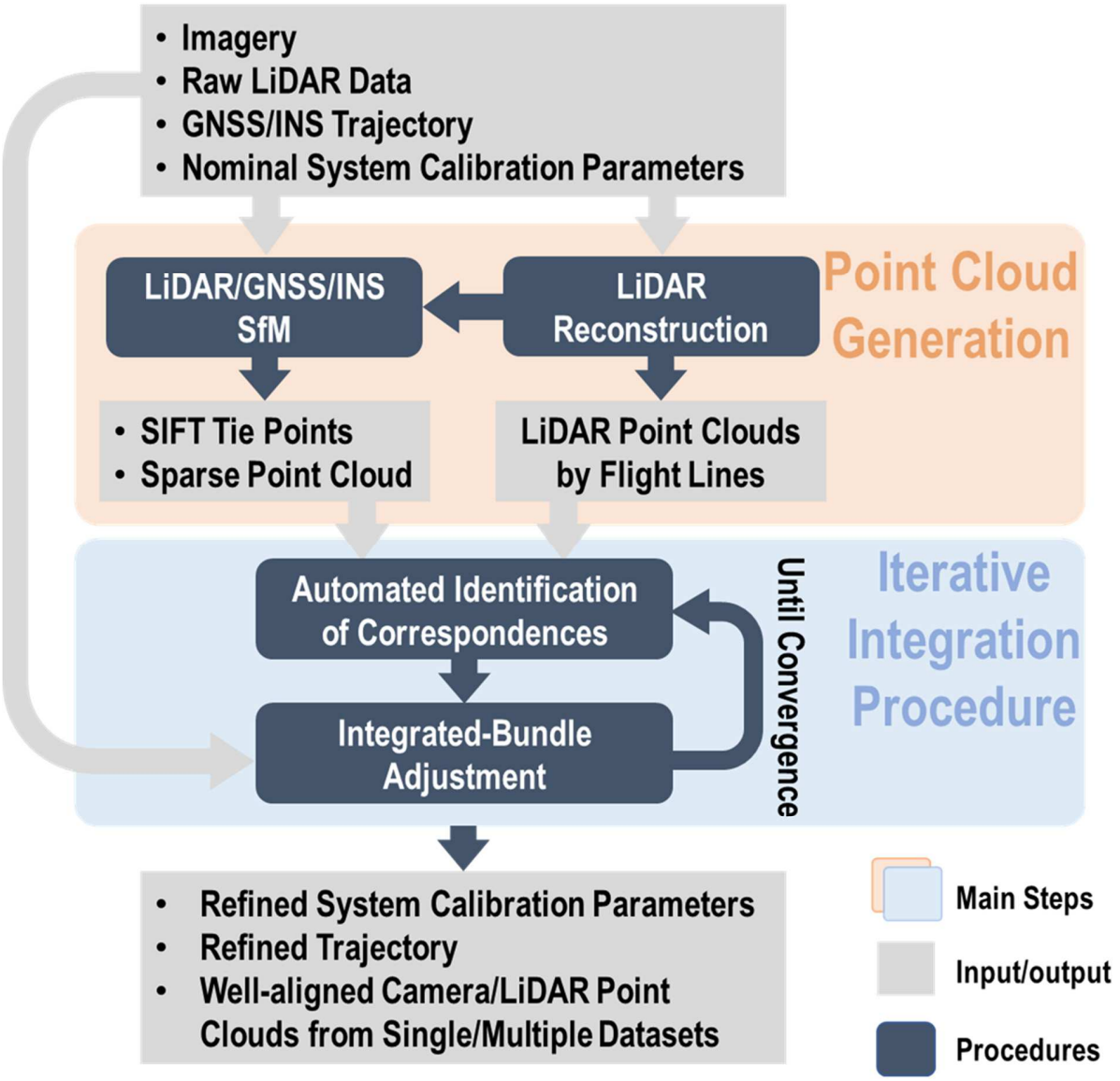


Figure 2. Workflow of the proposed tightly-coupled camera/LiDAR integration strategy.

Before discussing the details of the proposed LiDAR/GNSS/INS-assisted SfM, the mathematical models for deriving 3D geospatial information from LiDAR and camera sensors are presented. 3D reconstruction for LiDAR and image is based on the point positioning equation and modified collinearity equations, as represented by Equations 1 and 2, respectively. A schematic diagram of these two models is also illustrated in

Figure 3. In these equations, $r_I^{lu(t)}$, which is derived from raw LiDAR measurements at firing time (including range measurement and orientation of the laser beam relative to the laser unit reference frame), denotes the position of the laser beam footprint relative to the laser unit frame; $r_i^{c(t)}$ represents the image point i in the camera frame at exposure time; $\lambda(i,c,t)$ is an unknown scale factor for image point i captured by camera c at time t; $r_{b(t)}^m$ and $R_{b(t)}^m$ are the position and orientation information of the IMU body frame coordinate system relative to the mapping frame at firing/exposure time t; r_{lu}^b and R_{lu}^b represent the lever arm and boresight rotation matrix relating the laser unit system and the IMU body frame; r_c^b and R_c^b denote the lever arm and boresight rotation matrix relating the camera system and the IMU body frame; and r_I^m is the coordinates of object point I in the mapping frame. The modified collinearity equations (Equation 2) can be reformulated into Equation 3 where the image coordinates are represented as a function of the GNSS/INS information, camera IOP, camera mounting parameters, and ground coordinates of the object point. To eliminate the unknown scale factor – $\lambda(i,c,t)$ – from Equation 3, the first and second rows are divided by the third one, and the image point coordinates (x_i, y_i) are be expressed as per Equation 4.

$$r_I^m = r_{b(t)}^m + R_{b(t)}^m r_{lu}^b + R_{b(t)}^m R_{lu}^b r_I^{lu(t)} \tag{1}$$

$$r_I^m = r_{b(t)}^m + R_{b(t)}^m r_c^b + \lambda(i,c,t) R_{b(t)}^m R_c^b \, r_i^{c(t)} \tag{2}$$

where in $r_i^{c(t)} = \begin{bmatrix} x_i - x_p - dist_{x_i} \\ y_i - y_p - dist_{y_i} \\ -c \end{bmatrix}$: x_p and y_p are the principal point coordinates of the used camera, c is the principal distance of the used camera, $dist_{x_i}$ and $dist_{y_i}$ are the distortions in the x and y directions for image point i.

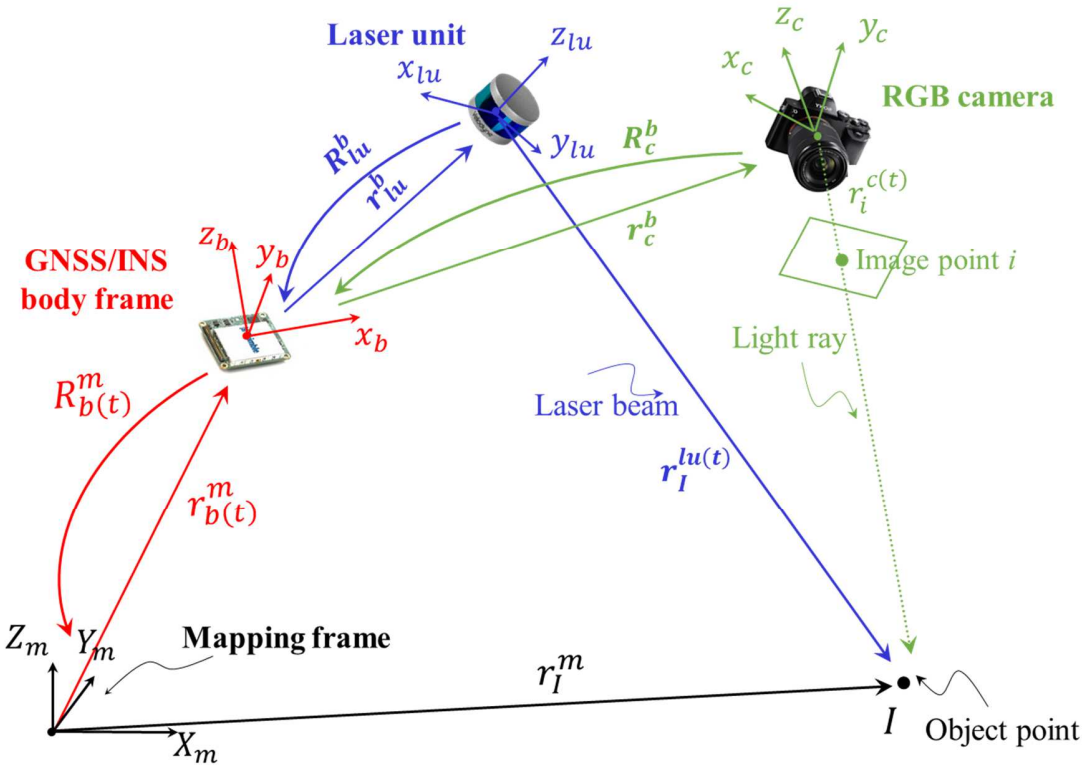


Figure 3. Schematic diagram of the point positioning equation and modified collinearity equations for LiDAR and camera units mounted on a GNSS/INS-assisted MMS.

$$r_i^{c(t)} = \frac{1}{\lambda(i,c,t)} \left[R_b^c R_m^{b(t)} \left(r_I^m - r_{b(t)}^m - R_{b(t)}^m r_c^b \right) \right] = \frac{1}{\lambda(i,c,t)} \begin{bmatrix} N_x \\ N_y \\ D \end{bmatrix} \tag{3}$$

$$x_i = -c \frac{N_x}{D} + x_p + dist_{x_i} \tag{4a}$$

$$y_i = -c \frac{N_y}{D} + y_p + dist_{y_i} \tag{4b}$$

As shown in Figure 2, the proposed framework starts with LiDAR reconstruction. A LiDAR point is directly reconstructed through the point positioning equation shown in Equation 1. In this study, flight lines are automatically extracted from the trajectory based on the heading angle. LiDAR point clouds coming from individual flight lines are then generated and used for extracting conjugate planar patches in the next stage. Moreover, the generated LiDAR point cloud are adopted in the proposed SfM strategy to provide approximate elevation information.

2.1. LiDAR/GNSS/INS-assisted SfM

The proposed image-based point cloud reconstruction strategy is adapted from the fully GNSS/INS-assisted SfM proposed in our previous work (Hasheminasab et al., 2020). In the GNSS/INS-assisted SfM, local features are first detected and characterized through descriptors using the scale invariant feature transform (SIFT) algorithm (Lowe, 2004). Rather than conducting exhaustive search among the feature descriptors within the images of a stereo-pair, GNSS/INS information and an average ground height are exploited to reduce the matching search space. More specifically, given a feature in the left image, an approximate location of its conjugate feature in the right image is estimated through a forward-backward projection strategy using i) an average ground elevation and ii) EOP derived from trajectory information and camera mounting parameters. This procedure is illustrated in Figure 4a. However, using a single ground

elevation makes the approach not suitable for applications where the terrain shows significant height variations.

In this study, a relatively sparse LiDAR point cloud is incorporated in the GNSS/INS-assisted SfM to provide approximate elevation information for each extracted image feature. This strategy – hereafter referred to as LiDAR/GNSS/INS-assisted SfM – starts with feature detection and characterization using the SIFT algorithm. Next, as shown in Figure 4b, for each SIFT feature extracted in the left image, its approximate 3D coordinates are estimated by finding the intersection between the corresponding image light ray and the visible surface defined by the available sparse LiDAR data. This is conducted by an octree-based ray tracing algorithm similar to the one proposed by Revelles et al. (2000). Then, the predicted point in the right image is derived through back-projection, and used to define a search window with a predefined size. Only the SIFT points within the search window in the right image will be used for feature matching. Figure 4c depicts for a feature in the left image (represented by blue solid circle), the estimated location of the conjugate feature using an average ground elevation (represented by blue cross) as well as LiDAR data (represented by blue triangle). It can be seen from this figure that the latter is more accurate than the former. This suggests that the probability of false matchings can be reduced by utilizing LiDAR point cloud.

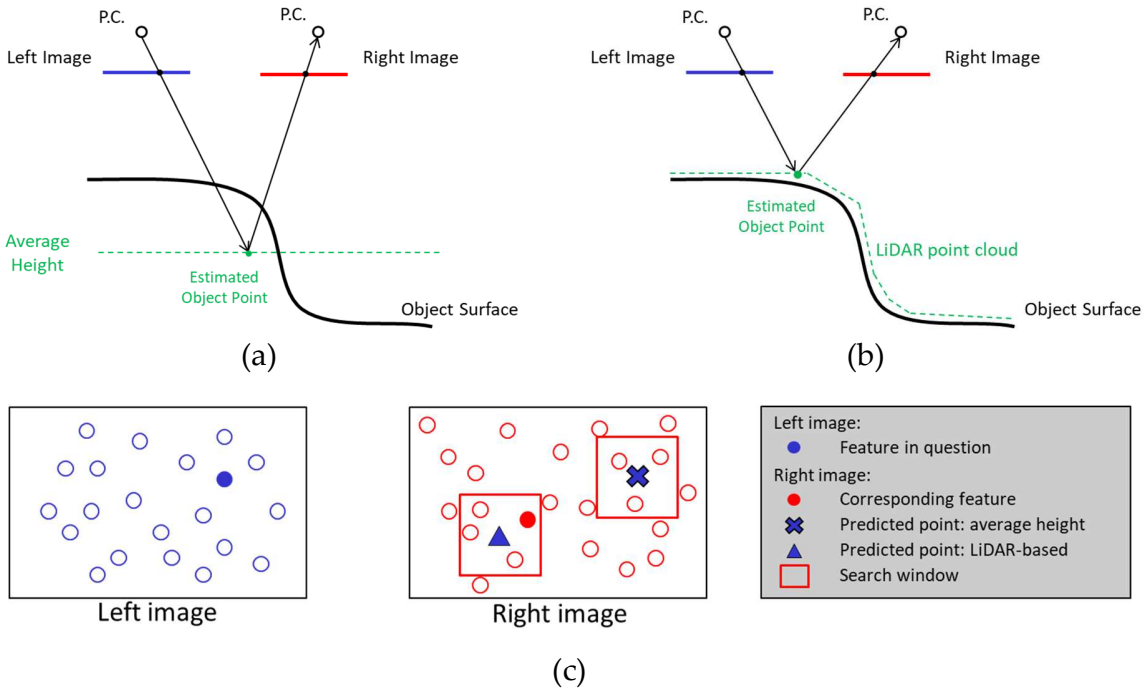


Figure 4. Illustration of forward-backward projection strategy using (a) average height and (b) LiDAR data, as well as (c) the selected feature in left image and its predicted locations using the average height and LiDAR data in right image.

Once matches are established in the stereo-pair, outliers are detected and removed using an iterative five-point algorithm based on epipolar geometry (He and Habib., 2016). In the next step, SIFT-based features are tracked among all involved imagery and then three outlier removal procedures – intersection-based, LiDAR-based, and iterative BA-based approaches – are adopted sequentially. First, the intersection-based outlier removal is implemented using a RANdom SAmple Consensus (RANSAC) procedure (Hasheminasab et al., 2020) to remove some matching outliers and derive approximate 3D coordinates of SIFT matches. Then, object points with large distance from their closest LiDAR points are detected and removed as further outliers. Finally, the iterative GNSS/INS-assisted BA procedure is conducted to remove more matching outliers and

derive an image-based, sparse point cloud. In the BA, the GNSS/INS trajectory is used as prior information and camera boresight angles are refined. After each iteration of the BA, outliers are removed through a chi-squared test with a predefined significance level on the residuals in the x and y coordinates of the image feature. The iterative procedure stops when the root-mean-square error (RMSE) of the back-projection error is smaller than a predefined threshold.

In summary, the proposed SfM strategy is implemented in the following steps: down-sampled LiDAR point cloud generation, SIFT feature extraction, height estimation of each extracted feature through ray tracing, feature matching (including the iterative five-point algorithm), feature tracking, RANSAC-based and LiDAR-based outlier removal, and iterative BA for sparse point cloud generation while removing more matching outliers. It is worth mentioning that the quality of established features in the proposed SfM is not sensitive to the accuracy of system calibration parameters. Such parameters are only utilized to restrict the search space and derive height estimates. As a result, the established SIFT tie features are not significantly affected by the system calibration parameters.

2.2. Automated Identification of Camera/LiDAR Feature Correspondences

Now that image-based point cloud (using the developed LiDAR/GNSS/INS-assisted SfM) and LiDAR point clouds from different flight lines (using point positioning equation) are generated while using nominal system calibration parameters,

image/LiDAR feature correspondences are established. The resultant image-based object points are used as seed points to identify corresponding planar patches in all LiDAR tracks through the iterative plane fitting algorithm introduced in Zhou et al. (2020). In the first step, a distance-based down-sampling is performed on the generated image-based sparse point cloud using a predefined distance threshold. This results in a balanced and uniformly distributed 3D object points which will improve the performance of the BA procedure.

The conceptual basis of identifying the planar patch through an iterative plane fitting is illustrated in Figure 5. For an image-based object point, its closest LiDAR point in a given track is first identified. If the distance between these two points is smaller than a predefined threshold, a spherical region centered at the selected LiDAR point with a predefined radius is created. LiDAR points within this spherical region are used to determine whether a planar patch exists or not through Principal Component Analysis (PCA) (Wold et al., 1987). If a planar patch exists, the iterative plane fitting process is conducted. In the iterative procedure, the plane derived in a given iteration is used to remove outlier points (based on their normal distance to the plane), followed by a subsequent plane fitting through the remaining points. This process will terminate when no more outliers are detected. The LiDAR points retained through this process, denoted as LiDAR inliers, constitute the planar patch. The derived planar patch is considered to be valid if the following criteria are met concurrently:

- The number of LiDAR inliers is larger than a predefined value,
- The ratio between LiDAR inliers and LiDAR points within the spherical region is larger than a predefined ratio, and
- The RMSE of normal distances for all LiDAR inliers to the best-fitting plane is smaller than a predefined maximum RMSE value.

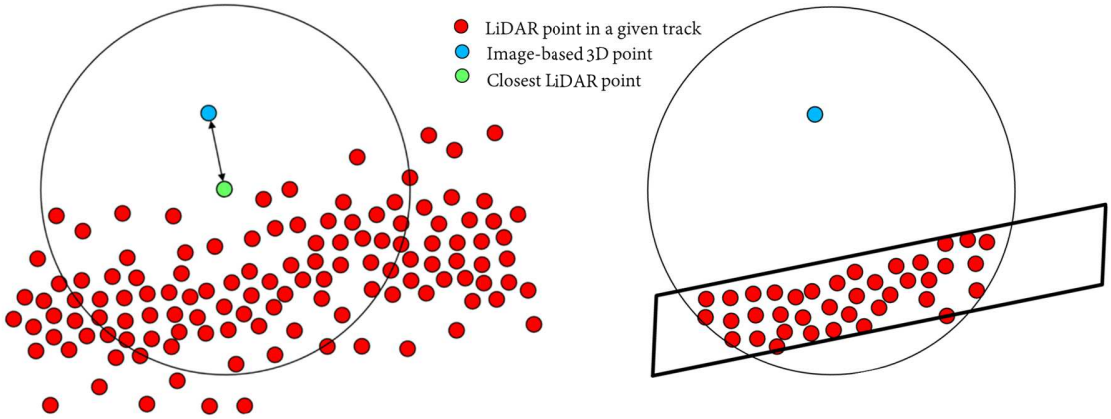


Figure 5. Schematic illustration of deriving the corresponding planar patch for an image-based 3D point from a LiDAR point cloud in a given track: (a) image-based 3D point (blue), LiDAR points (red), closest LiDAR point (green), and corresponding spherical region for extracting neighboring points and (b) LiDAR points retained through iterative plane fitting and derived best-fitting plane (adapted from Zhou et al., 2020).

For a given image-based 3D point, the iterative plane fitting process is conducted to derive its corresponding planar patches from all LiDAR tracks. An image-LiDAR correspondence primitive is derived if at least one point-to-planar patch correspondence is established. Such primitive consists of the respective SIFT tie features and LiDAR points constituting corresponding planar patches. After conducting the

automated identification of correspondences, derived primitives will be used in the BA for parameter refinement. As mentioned in Section 1, the proposed camera/LiDAR integration can also incorporate GCPs and control point clouds (e.g., LiDAR point cloud from a well-calibrated system). In case of GCPs availability, they are used as seed points (similar to what is conducted for image-based object points) to identify corresponding LiDAR planar patches. Manual measurements of GCPs in the visible images are considered as corresponding image tie points for that primitive. Also, when control point clouds are available, image-based seed points are used to identify corresponding planar patches in all control point clouds through the iterative plane fitting algorithm.

2.3. Integrated-Bundle Adjustment

Now that image/LiDAR correspondences are established, an integrated-bundle adjustment procedure is performed for deriving well-aligned image-based and LiDAR point clouds while refining system calibration parameters and trajectory information. The integrated-BA is carried out by simultaneously: i) enforcing the best intersection of conjugate light rays corresponding to SIFT-based tie points in overlapping images, ii) minimizing the normal distance among corresponding LiDAR patches extracted from different flight lines, iii) minimizing the normal distance from image-based object points and their corresponding LiDAR planar patches, iv) ensuring the intersection of conjugate light rays and LiDAR patches to be as close as possible to the corresponding GCP (in case of GCP availability), and v) minimizing the distance between image-based

object points/LiDAR planar patches to their corresponding planar patches from control point clouds (in case of control point clouds availability). Theoretically, in a well-calibrated system with accurate trajectory information, the image-based object points and corresponding LiDAR planar patches from different flight lines should be co-aligned along a planar patch. This planar patch – hereafter denoted as adjusted planar patch – can be represented by 3D coordinates of the image-based object point I, $r_I^m = [X_I, Y_I, Z_I]^T$, and a normal vector, $\vec{\boldsymbol{n}} = [n_x, n_y, n_z]^T$ (see Figure 6). Since the normal vector $\vec{\boldsymbol{n}}$ has two degrees of freedom, it is represented by two angles: α and β. In total, five parameters, (r_I^m, α, β), are introduced for each correspondence primitive. For a LiDAR point L lying on this adjusted planar patch, Equation 5 must be satisfied. The adopted mathematical models for the objective functions in the integrated-BA are presented below.

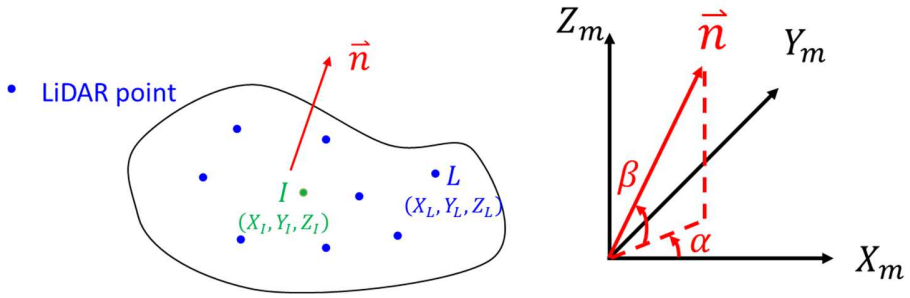


Figure 6. Representation of the adjusted planar patch according to a camera/LiDAR correspondence primitive (image-based object point denoted by I and LiDAR points denoted by L).

$$\vec{\boldsymbol{n}}^T \cdot \begin{bmatrix} X_L - X_I \\ Y_L - Y_I \\ Z_L - Z_I \end{bmatrix} = 0 \Rightarrow n_x X_L + n_y Y_L + n_z Z_L - \vec{\boldsymbol{n}}^T \cdot r_I^m = 0 \quad (5)$$

where: $[X_L, Y_L, Z_L]^T = r_L^m$ is the 3D coordinates of the LiDAR point L, $n_x = \cos(\alpha)\cos(\beta)$, $n_y = \sin(\alpha)\cos(\beta)$, and $n_z = \sin(\beta)$.

i) SIFT-based tie points: To ensure the best intersection of SIFT-based tie points, back-projection error, which is modeled using the modified collinearity equations (Equation 4), is adopted as a cost function for conjugate image points. The back-projection error of image point i in the x and y directions is a function of the image coordinates – (x_i, y_i), camera IOP, camera mounting parameters – r_c^b / R_c^b, position/orientation of the IMU body frame at the exposure time t_i – $r_{b(t_i)}^m / R_{b(t_i)}^m$, and 3D coordinates of the image-based object point I – r_I^m, as presented in Equation 6.

ii) Conjugate LiDAR planar patches: The cost function for LiDAR points is defined as their normal distances to the adjusted planar patch. To do so, the 3D coordinates of the LiDAR point, r_L^m, are first derived through the point positioning equation (Equation 1). Next, distance from the LiDAR point, L, to this plane can be computed by Equation 7. In summary, as presented in this equation, the residual term – normal distance to the adjusted planar patch – of a LiDAR point is a function of the raw LiDAR measurements implicitly present in $r_L^{lu}(t_L)$, LiDAR mounting parameters – r_{lu}^b / R_{lu}^b, position/orientation of the IMU body frame at the

384 firing time $t_L - r^m_{b(t_L)}/R^m_{b(t_L)}$, and the five parameters representing the adjusted
385 planar patch – (r_I^m, α, β).
386 iii) Image-based object point and corresponding LiDAR planar patches: The
387 constraints that image-based object point should coincide with the LiDAR planar
388 patches are implicitly considered in the residual terms for image tie points and
389 LiDAR points (Equations 6 and 7) – i.e., defining the adjusted planar patch to be as
390 close as possible to the image-based object point and LiDAR points. Since the
391 adjusted planar patch is defined by the image-based object point along with a
392 normal vector, minimizing the distance from LiDAR points to the adjusted planar
393 patch is equivalent to forcing the image-based object point to be aligned with the
394 LiDAR planar patches.
395 iv) Primitive derived from a GCP: Figure 7a shows an example of a primitive derived
396 from a GCP. As shown in the figure, prior information is provided for the three
397 unknowns – r_I^m – of the adjusted planar patch since the patch must pass through
398 the GCP. If image measurements of the GCP are available, both objective functions
399 relating to image and LiDAR observations (Equations 6 and 7) are applicable.
400 Otherwise, only residuals for LiDAR points – Equation 7 – are included in the BA
401 process. The prior information provided by the GCP is included through pseudo
402 observations – i.e., direct observations of the unknowns – of the three unknowns
403 (r_I^m).

v) Primitive with a planar patch from a control point cloud: An example of a primitive with a planar patch extracted from a control point cloud – hereafter denoted as control planar patch – is illustrated in Figure 7b. The control planar patch can enforce the location and orientation of the unknown adjusted planar patch. In other words, the adjusted planar patch should be co-aligned with the control planar patch (see Figure 7b). By estimating the parameters of the control planar patch, prior information for the five unknowns of the adjusted planar patch can be derived. In order to incorporate the prior information, pseudo observations of the unknowns are included in the BA process.

$$res_{Image} = f\left((x_i,\ y_i),\ camera\ IOP,\ r_c^b, R_c^b, r_{b(t_i)}^m, R_{b(t_i)}^m, r_I^m\right) \tag{6}$$

$$\begin{aligned} res_{LiDAR} = d &= \left|n_x X_L + n_y Y_L\ + n_z Z_L - \vec{\boldsymbol{n}}^{\,T} \cdot r_I^m\right| \\ &= f\left(r_L^{lu(t_L)}, r_{lu}^b, R_{lu}^b, r_{b(t_L)}^m, R_{b(t_L)}^m, r_I^m, \alpha, \beta\right) \end{aligned} \tag{7}$$

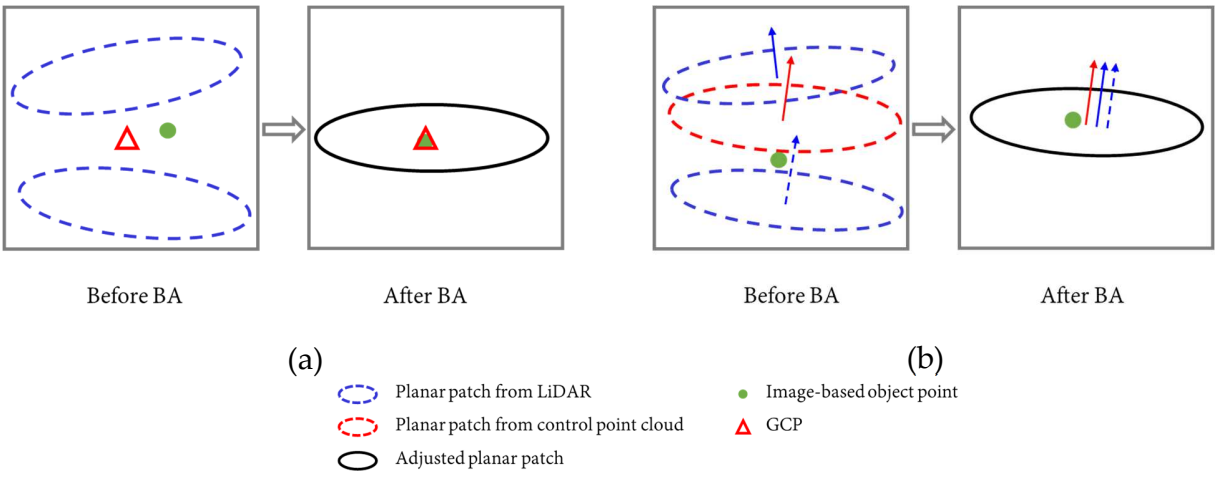


Figure 7. Schematic illustration of primitives and their corresponding adjusted planar patches when: (a) GCP or (b) control point cloud is available.

In conclusion, considering a primitive with n image tie points and m LiDAR points in all corresponding planar patches, $(2n + m)$ equations will be derived. The a-priori variances of the image and LiDAR observations are determined based on their expected accuracy. The former is selected according to the quality of established SIFT tie points in the SfM process while the latter depends on the accuracy of the hardware involved in the LiDAR reconstruction process. As shown in Equations 6 and 7, the observations include the coordinates of SIFT image points and LiDAR raw measurements while other parameters including system calibration parameters, adjusted planar patch parameters, and trajectory information can be estimated/refined in the integrated-BA procedure. One should note that while trajectory information at the camera exposure time, t_i, for all SIFT points in a given image remains the same, every LiDAR point has its unique time stamp, t_L, in the trajectory. Therefore, solving for the position and orientation of the IMU body frame at time stamps corresponding to each LiDAR point, t_L, is impractical. To overcome this issue, this study exploits reference points extracted from the GNSS/INS trajectory through which the position and orientation of the IMU body frame at any time stamp can be interpolated.

In general, data rate of the trajectory generated from the post-processing of raw GNSS/INS data is relatively high (e.g., 200 Hz). To reduce the number of position and orientation unknowns, reference points are first extracted using a lower data rate. In case that a p^{th} order polynomial is used, a position/orientation parameter $-\theta-$ of a

reference point l with timestamp t_l can be expressed by Equation 8. For an image/LiDAR observation with time stamp t_{obs}, q reference points with closest time stamps to t_{obs} are identified ($q \geq p+1$) to represent the parameter $\theta(obs)$ in question, as depicted in Figure 8. The position/orientation parameter θ for these q neighboring reference points should satisfy the p^{th} order polynomial presented in Equation 8, thus leading to Equation 9. To avoid possible numerical instabilities, the time stamp t is replaced by a relative time t' (i.e., relative time for the first reference point is set to zero – $t'_l = 0$) in Equation 9. Then, the coefficients of the p^{th} order polynomial for the parameter θ can be derived based on LSA principles (as long as $q \geq p+1$ is satisfied) as in Equation 10. Finally, the position/orientation parameter at time stamp t_{obs} for an image/LiDAR observation is derived by Equation 11. In other words, the position/orientation parameter of a given observation is a linear combination of that from its closest q successive reference points.

$$\theta(l) = \begin{bmatrix} 1 & t_l & t_l^2 & \cdots & t_l^p \end{bmatrix} \cdot \begin{bmatrix} C_\theta(0) \\ C_\theta(1) \\ C_\theta(2) \\ \vdots \\ C_\theta(p) \end{bmatrix} \tag{8}$$

where: $C_\theta(0), C_\theta(1), \ldots, C_\theta(p)$ are the coefficients of the p^{th} order polynomial for parameter θ.

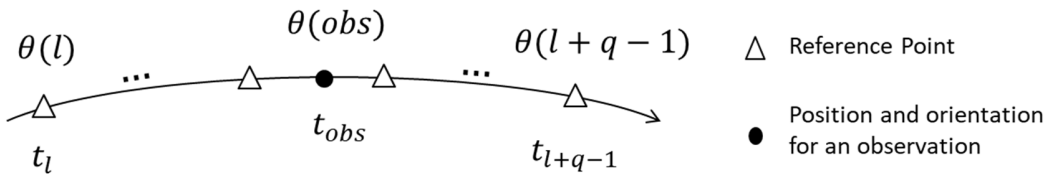


Figure 8. Identification of neighboring reference points for an observation with time stamp t_{obs} to derive its position/orientation parameter through interpolation.

449

$$\begin{bmatrix} \theta(l) \\ \theta(l+1) \\ \theta(l+2) \\ \vdots \\ \theta(l+q-1) \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & t'_l & t'^2_l & \cdots & t'^p_l \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t'_{l+q-1} & t'^2_{l+q-1} & \cdots & t'^p_{l+q-1} \end{bmatrix}}_{\boldsymbol{K}} \cdot \begin{bmatrix} C_\theta(0) \\ C_\theta(1) \\ C_\theta(2) \\ \vdots \\ C_\theta(p) \end{bmatrix} \tag{9}$$

450 where:

451 t' is the relative time, $t'_{l+j} = t_{l+j} - t_l$ $(j = 0, 1, 2, \ldots, q-1)$, and

452 K is a matrix defined by the relative time associated with the q reference points.

$$\begin{bmatrix} C_\theta(0) \\ C_\theta(1) \\ C_\theta(2) \\ \vdots \\ C_\theta(p) \end{bmatrix} = (\boldsymbol{K^T K})^{-1} \boldsymbol{K^T} \cdot \begin{bmatrix} \theta(l) \\ \theta(l+1) \\ \theta(l+2) \\ \vdots \\ \theta(l+q-1) \end{bmatrix} \tag{10}$$

$$\begin{aligned} \theta(obs) &= \begin{bmatrix} 1 & t'_{obs} & t'^2_{obs} & \cdots & t'^p_{obs} \end{bmatrix} \begin{bmatrix} C_\theta(0) \\ C_\theta(1) \\ C_\theta(2) \\ \vdots \\ C_\theta(p) \end{bmatrix} \\ &= \begin{bmatrix} 1 & t'_{obs} & t'^2_{obs} & \cdots & t'^p_{obs} \end{bmatrix} \cdot (\boldsymbol{K^T K})^{-1} \boldsymbol{K^T} \cdot \begin{bmatrix} \theta(l) \\ \theta(l+1) \\ \theta(l+2) \\ \vdots \\ \theta(l+q-1) \end{bmatrix} \end{aligned} \tag{11}$$

453 where: $t'_{obs} = t_{obs} - t_l$.

By interpolating trajectory information related to the observations, only the GNSS/INS position and orientation parameters of the selected reference points are regarded as unknowns in the integrated-BA model. To introduce the prior information provided by the GNSS/INS unit, pseudo observations are adopted. The variances of the position and orientation observations are determined according to the accuracy specifications of the utilized GNSS/INS unit. Six unknowns and six additional observations are introduced for each reference point along the trajectory.

2.4. Workflow of the Proposed Camera/LiDAR Integration Strategy

The proposed camera/LiDAR integration framework starts with LiDAR/image-based point cloud reconstruction using the point positioning equation and LiDAR/GNSS/INS-assisted SfM, respectively. In the developed SfM pipeline, other than image-based sparse point cloud, SIFT tie points are also established. Then, the identification of camera/LiDAR correspondences and integrated-bundle adjustment steps are conducted iteratively (see Figure 2). It is worth mentioning that the LiDAR/GNSS/INS-assisted SfM is not included in the iterative process since the established SIFT tie features are not significantly affected by the quality of utilized system calibration parameters. As a result, once refined system calibration parameters are available, updated image-based sparse point cloud can be derived using the SIFT tie points established in the first step.

In the iterative process mentioned above, system calibration parameters are refined in each iteration. As for the trajectory, initial values provided by GNSS/INS unit will be used throughout the iterations in the integrated-BA. As shown in the framework (Figure 2), before starting a new iteration, image-based and LiDAR point clouds need to be updated using the refined system calibration parameters. For each LiDAR point, its 3D coordinates will be computed using the initial trajectory and refined LiDAR mounting parameters. The image-based sparse point cloud is updated through a GNSS/INS-assisted BA using the refined IOP. In this BA, boresight angles and trajectory information for each image are refined to ensure the best intersection of light rays in the object space. The reason behind utilizing GNSS/INS-assisted BA instead of multi-light ray intersection to update the image-based point cloud is that the former can produce better estimates of the 3D coordinates. By doing this, more correspondences can be established in next iteration.

The iterative process terminates when the changes in the system calibration parameters are smaller than a predefined threshold. Other than the estimated parameters and refined trajectory derived from the last iteration, final LiDAR and image-based point clouds are also generated. The former is reconstructed using point positioning equation with the refined trajectory and estimated LiDAR mounting parameters, and the latter is conducted by multi-light ray intersection. More specifically, refined EOP of each image are computed by the estimated camera

493 mounting parameters and refined trajectory information. Then, image-based sparse
494 point cloud is derived using the refined EOP and established SIFT tie points through
495 multi-light ray intersection. One should note that in the first few iterations, due to the
496 inaccurate system calibration parameters, the alignment among image-based object
497 points and LiDAR point clouds from different flight lines is relatively poor. Therefore,
498 more relaxed thresholds for establishing corresponding features are used. Stricter
499 thresholds are used as the alignment gets better throughout the iterations. In addition,
500 to improve the computational efficiency, fewer number of image-based/LiDAR points
501 are used for correspondence identification in the early stages of the integration process.
502 As mentioned in Section 2.2, the distance-based down-sampling is performed on the
503 image-based sparse point. In terms of LiDAR data, number of LiDAR points can be
504 reduced by limiting the reconstruction to every few firings. For a laser unit with
505 multiple channels, down-sampling can also be conducted by reconstructing a subset of
506 the channels.

507 Moreover, the proposed integration workflow can be adopted for multiple systems
508 at the same time to produce well-aligned point clouds from different sensors and
509 multiple systems. To do so, two strategies can be adopted: i) SfM is simultaneously
510 conducted on images from all involved cameras to generate one set of sparse point
511 cloud, or ii) images from different cameras are processed individually to generate a
512 sparse point cloud for each sensor. The main difference between the two approaches is

513 that the first strategy integrates images from different cameras directly by finding
514 corresponding features while the second one integrates each camera to LiDAR data
515 separately. Hence, the former first ensures the alignment among all cameras, and then
516 integrates the combined image data to LiDAR point clouds. The latter indirectly ensures
517 the alignment among image-based point clouds through integrating them with LiDAR
518 data individually. Considering that feature matching for images from different datasets
519 is challenging (e.g., different lighting conditions, scale), second strategy is adopted.
520 Once the sparse point clouds are generated, for an image-based object point, its
521 corresponding planar patches from all LiDAR tracks and all systems are identified. This
522 process is repeated for all image-based object points from these multiple systems.
523

524 **3. Specifications of Data Acquisition Systems and Experimental Datasets**

525 In this section, the platforms and sensors used in this study are introduced. Then,
526 the acquired datasets that are used to evaluate the performance of the proposed
527 camera/LiDAR integration workflow are described.

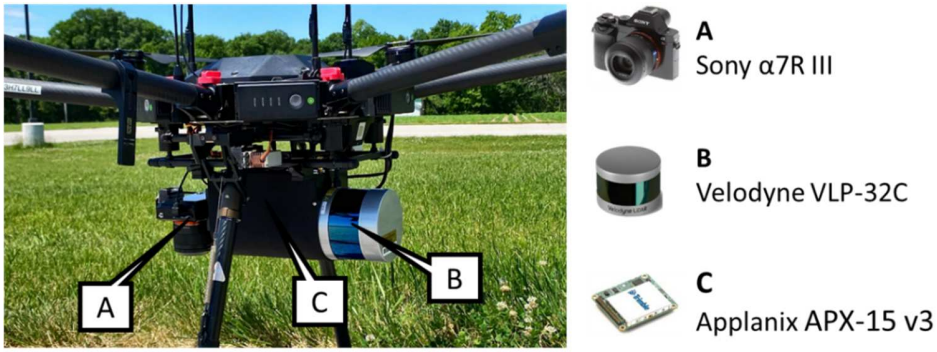
528 **3.1. Data Acquisition Systems**

529 In this study, two custom-built UAV mobile mapping systems, as shown in Figure 9,
530 are deployed for imagery and LiDAR data acquisition. Both systems consist of a Dà-
531 Jiāng Innovations (DJI) Matrice 600 Pro carrying an RGB camera, a LiDAR sensor, and
532 an

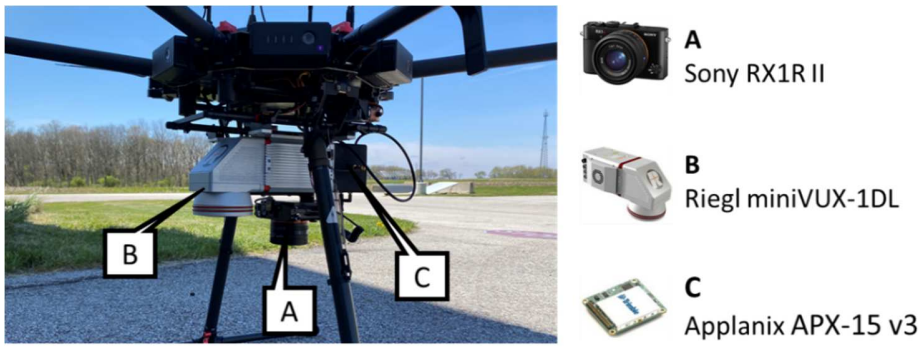
Applanix APX-15 UAV v3 GNSS/INS unit. The RGB camera, LiDAR, and GNSS/INS position and orientation system are rigidly fixed relative to one another. Direct geo-refencing information, i.e., the position and orientation of the system, is provided by the APX-15 v3 unit at 200 Hz data rate. After post-processing the GNSS/INS data, the expected positional accuracy is ±2-5 cm, and the accuracy for pitch/roll and heading is ±0.025° and ±0.08°, respectively (Applanix, 2021).

The first UAV system, hereafter denoted as "*VLP-system*", employed a Sony α7R III camera and a Velodyne VLP-32C, as shown in Figure 9a. The Sony α7R III is a 42-megapixel camera with a 7952 × 5304 complementary metal oxide semiconductor (CMOS) array, 4.5 μm pixel size, and a lens with 35 mm nominal focal length (Sony, 2021a). The Velodyne VLP-32C consists of 32 channels. This sensor fires approximately 600,000 pulses per second with a 360° horizontal field of view (FOV) and a 40° vertical field of view (+15° to -25° from the horizon – i.e., direction perpendicular to the axis of rotation). The maximum measurement range is 200 m with a ±3 cm range accuracy (Velodyne, 2021). The rotational axis of the VLP-32C is along the flying direction of the UAV. In this study, a LiDAR point is reconstructed only when the direction of corresponding laser beam is less than ±70° from nadir. As a result, at most 44% of the 600,000 points per second can be reconstructed, assuming that the measurement range is within 200 m.

552 The second UAV system, hereafter denoted as "*Riegl-system*", carries a Sony RX1R II
553 camera and a Riegl miniVUX-1DL laser scanner (see Figure 9b). The Sony RX1R II
554 camera has identical specifications as the Sony α7R III in terms of image resolution,
555 pixel size, and nominal focal length (Sony, 2021b). The Riegl miniVUX-1DL adopts a
556 rotating wedge prism to generate a circular scan pattern with 23° off-nadir scan angle
557 (i.e., a FOV of 46°). This laser unit can conduct 100,000 range measurements per second.
558 The maximum measurement range is from 140-260 m depending on the incidence angle
559 with a ±1.5 cm range accuracy (Riegl, 2021). Different from the Velodyne VLP-32C,
560 theoretically all range measurements can be reconstructed due to the scanning
561 characteristics of the Riegl miniVUX-1DL.



(a)



(b)

Figure 9. The UAV-based mobile mapping systems and onboard sensors used in this study: (a) *VLP-system* and (b) *Riegl-system*.

The cameras for both systems are triggered at a frame interval of 1.5 s, and an event marker for each image is recorded in the GNSS/INS trajectory through a feedback signal from the camera's flash hotshot to the GNSS/INS unit at the time of exposure (Elbahnasawy, 2018). Initial values of camera IOP are estimated through an indoor calibration procedure (Habib and Morgan, 2005). In this study, camera IOP include principal distance (c) and distortion parameters (K_1, K_2, P_1, P_2). Nominal boresight angles of each sensor to the IMU body frame are derived based on the system setup, while initial values for lever arm components are established through tape measurements. The initial system calibration parameters for these two systems are listed in Table 1, where ($\Delta\omega_c, \Delta\phi_c, \Delta\kappa_c$) and ($\Delta X_c, \Delta Y_c, \Delta Z_c$) denote the boresight angles and lever-arm components relating the camera coordinate system relative to the GNSS/INS body frame, respectively. Similarly, LiDAR mounting parameters are denoted as ($\Delta\omega_{lu}, \Delta\phi_{lu}, \Delta\kappa_{lu}$) and ($\Delta X_{lu}, \Delta Y_{lu}, \Delta Z_{lu}$).

Table 1. Initial system calibration parameters for *VLP-system* and *Riegl-system*.

System	$\boldsymbol{c}$ (pixel)	$\boldsymbol{K1}$ (10^{-10} pixel^{-2})	$\boldsymbol{K2}$ (10^{-17} pixel^{-4})	$\boldsymbol{P1}$ (10^{-7} pixel^{-1})	$\boldsymbol{P2}$ (10^{-8} pixel^{-1})	$\boldsymbol{\Delta\omega_c}$ (°)	$\boldsymbol{\Delta\phi_c}$ (°)	$\boldsymbol{\Delta\kappa_c}$ (°)
VLP	8030.45	8.41	-5.06	1.36	-8.76	180	0	-90
Riegl	7558.19	-11.42	-2.94	-1.28	-3.69	180	0	-90

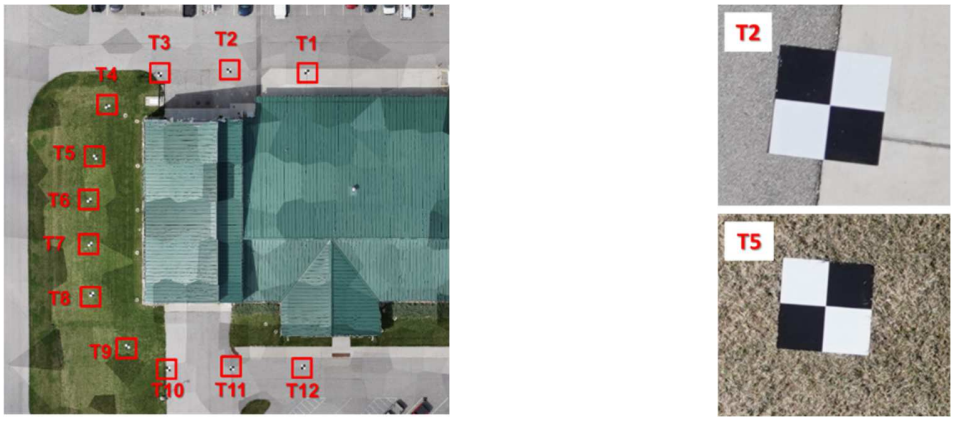
Systems	$\boldsymbol{\Delta X_c}$ (cm)	$\boldsymbol{\Delta Y_c}$ (cm)	$\boldsymbol{\Delta Z_c}$ (cm)	$\boldsymbol{\Delta\omega_{lu}}$ (°)	$\boldsymbol{\Delta\phi_{lu}}$ (°)	$\boldsymbol{\Delta\kappa_{lu}}$ (°)	$\boldsymbol{\Delta X_{lu}}$ (cm)	$\boldsymbol{\Delta Y_{lu}}$ (cm)	$\boldsymbol{\Delta Z_{lu}}$ (cm)

VLP	10	0	5	0	0	0	-10	0	5
Riegl	8	0	8	-90	0	-90	22	0	4

576

577 **3.2. Dataset Description**

578 In this study, five datasets were collected over a study site to evaluate the
579 performance of the proposed camera/LiDAR integration process. The study site is
580 located at the Agronomy Center for Research and Education (ACRE) in Purdue
581 University with different geomorphic features, i.e., grass, pavement, and building roof,
582 as shown in Figure 10. Twelve highly reflective checkerboard targets, which can be
583 identified in both imagery and LiDAR data, were deployed in the study site (red boxes
584 in Figure 10a). The checkerboard target is a 0.75 m by 0.75 m galvanized steel plate with
585 retro reflective vinyl on it. These targets are used as check points (CPs) to evaluate the
586 absolute accuracy of the integration results. Coordinates of the centers of these
587 checkerboard targets were determined using a real-time kinematic (RTK)-GNSS survey
588 with a Trimble R10 GNSS receiver. Considering the distance between the rover receiver
589 and the GNSS base-station – i.e., 6 km for the study site – the expected horizontal and
590 vertical accuracy is in the range of ±2 to ±3 cm and ±3 to ±4 cm, respectively (Trimble,
591 2021).



(**a**) (**b**)

Figure 10. (a) Study site and with red boxes highlighting the deployed checkboard targets and (b) sample images of the targets.

592 A summary of utilized systems and flight configurations of the collected
593 experimental datasets is presented in Table 2. The three datasets that were collected on
594 different dates using the *VLP-system* are denoted as A-1, A-2, and A-3. The proposed
595 integration process will be conducted individually on these datasets. The remaining
596 two datasets, denoted as B-1 and B-2, were acquired on the same date using *VLP-system*
597 and *Riegl-system*, respectively. These two datasets will be processed simultaneously to
598 validate the capability of the proposed strategy in integrating camera and LiDAR data
599 from multiple systems. A single mission plan, which consists of twelve east–west flight
600 lines and three north–south flight lines, was designed using the DJI GS Pro software for
601 all datasets. The duration of the mission was approximately 6 minutes. Figure 11 shows
602 a top view of the flight trajectory (colored by time) superimposed on an orthophoto for
603 the A-1 dataset. Examples of reconstructed LiDAR point clouds for the B-1 and B-2
604 datasets, as well as a sample profile (3 m by 0.2 m) extracted from a single LiDAR track

605 in the two point clouds are shown in Figure 12. Here, the maximum reconstruction
606 range is set to 100 m for both LiDAR sensors. It can be seen from Figure 12a and Figure
607 12b that VLP-32C with approximately 42 million points covers larger area compared to
608 the Riegl laser unit with approximately 26 million points. However, by closer inspection
609 of Figure 12c, one can note that the noise level in the profile extracted from point cloud
610 produced by the Riegl unit is lower than that in VLP-32C point cloud. The variation in
611 the noise level can be attributed to the different range accuracy of the two laser units.

612 Table 2. System and flight configuration for the collected experimental datasets.

	Dataset Name		A-1	A-2	A-3	B-1	B-2
	Date		20200304	20200317	20200501	20210425	
	System			*VLP*		*VLP*	*Riegl*
	Flying Height (m)			41		41	
	Ground Speed (m/s)			4.0		4.0	
	Lateral Distance (m)	East-West Flight Lines		6		6	
		North-South Flight Lines		20		20	
Image-specific	**Ground Sampling Distance (cm)**			0.6		0.6	
	Overlap/ Side-Lap (%)	East-West Flight Lines		83/89		83/89	
		North-South Flight Lines		83/53		83/53	
	Number of Captured Images		209	210	209	211	211
LiDAR-specific	**LiDAR Side-Lap (%)**	East-West Flight Lines		95		95	82
		North-South Flight Lines		83		83	43
	Number of LiDAR Points		~43M	~44M	~42M	~42M	~26M
	Covered Area (m²)			~24K		~24K	~14K

613

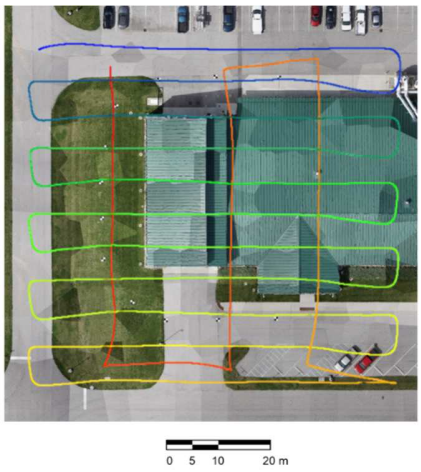


Figure 11. Top view of the flight trajectory (colored by time) superimposed on an orthophoto of the study site for the A-1 dataset.

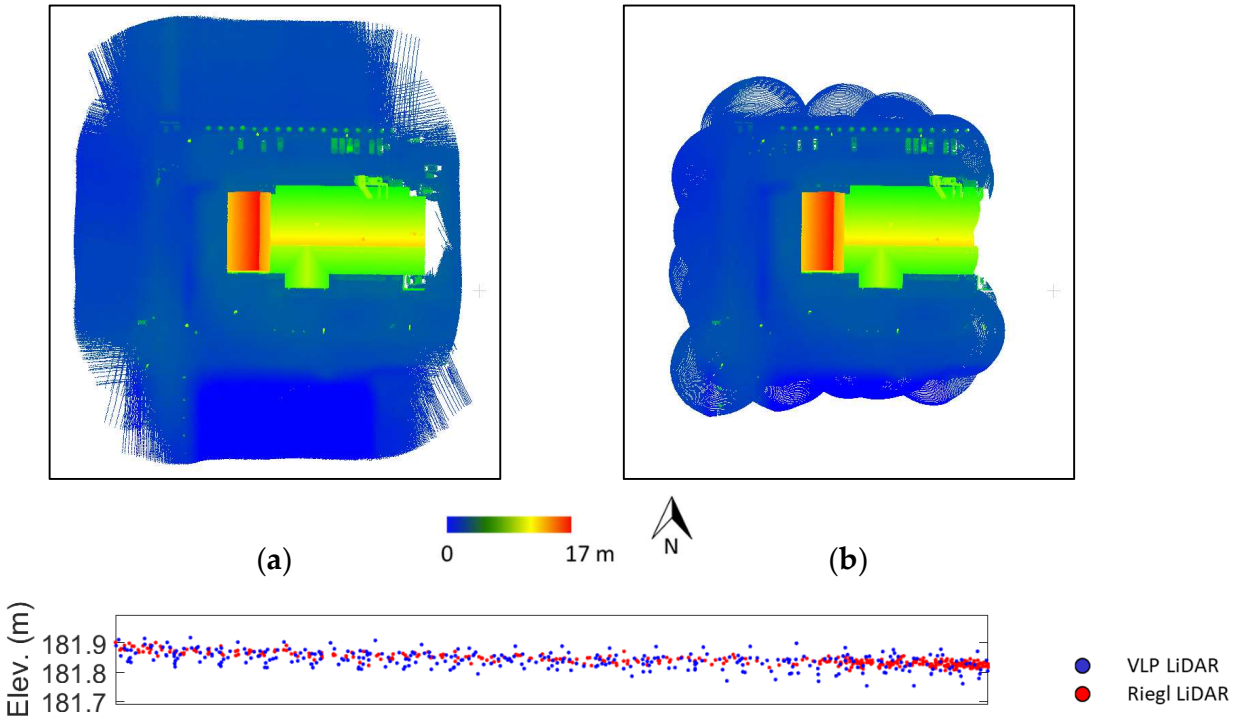


Figure 12. Generated LiDAR point clouds (colored by height) for (a) the B-1 dataset with ~42 million points and (b) the B-2 dataset with ~26 million points, as well as (c) side-view of a sample profile from each dataset.

4. Experimental Results and Discussion

In this section, the performance of the LiDAR/GNSS/INS-assisted SfM is first evaluated through a comparison with the GNSS/INS-assisted SfM using an average height. The capability of the proposed camera/LiDAR integration framework is then verified for the five datasets by qualitatively and quantitively analyzing the relative/ absolute accuracy of the image-based and LiDAR point clouds.

4.1. Performance Evaluation of LiDAR/GNSS/INS-assisted SfM

In this subsection, the performance of the proposed LiDAR/GNSS/INS-assisted SfM, hereafter denoted as "*LiDAR-SfM*", is evaluated and compared to the approach without using LiDAR, hereafter denoted as "*Average-Height-SfM*". The latter is implemented in the following steps: SIFT feature extraction, feature matching, feature tracking, RANSAC-based outlier removal, and iterative BA-based outlier removal/sparse point cloud generation – i.e., LiDAR reconstruction, height estimation for SIFT features, and LiDAR-based outlier removal are excluded in the *Average-Height-SfM*. The utilized threshold values in the SfM strategies are listed in Table 3. Using the 0.6 m distance threshold reported in this table for LiDAR point cloud down-sampling, number of LiDAR points decreases from ~42 million to ~44 thousand and from ~26 million to ~23 thousand for the B-1 and B-2 datasets, respectively.

Table 3. Selected threshold values for the SfM strategies.

Threshold Description	Threshold Value
Distance threshold for LiDAR point cloud down-sampling	0.6 m
Search window size in feature matching	500 × 500 pixels
Maximum distance in the LiDAR-based outlier removal	5 m
Significance level for the chi-squared test in the iterative BA-based outlier removal	2.5%
Maximum RMSE of back-projection errors to stop the iterative BA-based outlier removal	2.5 pixel

In addition to the above-mentioned parameters, in the implemented iterative GNSS/INS-assisted BA, a-priori variance for the image coordinate measurements (i.e., SIFT features) is set to (7 pixels)^2 based on their expected accuracy. In terms of the pseudo observations for GNSS/INS information, a-priori variance values for the position, roll/pitch angles, and heading angle are set to be (3 cm)^2, (0.025°)^2, and (0.08°)^2, respectively. In this process, boresight angles ($\Delta\omega_c, \Delta\phi_c, \Delta\kappa_c$) relating the camera coordinate system to the GNSS/INS body frame are estimated in the BA.

The comparison between *LiDAR-SfM* and *Average-Height-SfM* is conducted using the following measures:

- Square root of a-posteriori variance factor ($\hat{\sigma}_0$) and standard deviations of the estimated parameters from BA: The former indicates whether appropriate a-priori variances are set to the observations, while the latter reveals the quality of estimated parameters;
- RMSE value of the back-projection errors after BA: This value illustrates the quality of established conjugate features using the refined boresight angles;
- Number of images with insufficient tracked SIFT features: Number of images in which less than 20 SIFT tie points have been tracked;
- Number of reconstructed object points and tracked SIFT features: This measure represents the ability of the SfM framework for identifying conjugate features in overlapping images and establishing object points;

- Processing time: Individual processing time for each step of the two SfM frameworks are listed to evaluate the computational efficiency.

The two frameworks are adopted for the five datasets introduced in Section 3, and the derived square root of a-posteriori variance factor ($\hat{\sigma}_0$), estimated boresight angles along with their standard deviations are presented in Table 4. The estimated $\hat{\sigma}_0$ values for both strategies are in the range of 0.3-0.5, indicating that the assigned a-priori variances for the SIFT tie points and trajectory information are conservative. It can be observed from Table 4 that boresight pitch and heading angles ($\Delta\omega_c$ and $\Delta\kappa_c$) from the two SfM strategies are similar, while differences in boresight roll angle ($\Delta\phi_c$) are around 0.1°, which leads to 7 cm different on the object space along the flying direction at 41 m flying height. This can be caused by the correlation among $\Delta\phi_c$ and trajectory information, as the overlap along the flying direction is usually small (in this study, X axis of the IMU body frame is along the flying direction). In terms of the standard deviations for the estimated boresight angles, the values from both strategies are similar.

669 Table 4. Derived square root of a-posteriori variance factor, estimated boresight angles
670 and their standard deviations from the two SfM strategies for the five datasets.

Dataset	SfM Strategy	$\widehat{\sigma}_0$	$\Delta\omega_c$ (°)	$\Delta\phi_c$ (°)	$\Delta\kappa_c$ (°)
A-1	*LiDAR*	0.41	180.40 ± 0.001	0.72 ± 0.001	-89.54 ± 0.003
	Average-Height	0.42	180.35 ± 0.002	0.63 ± 0.002	-89.56 ± 0.004
A-2	*LiDAR*	0.35	180.04 ± 0.001	0.75 ± 0.001	-89.57 ± 0.003
	Average-Height	0.48	180.02 ± 0.002	0.59 ± 0.002	-89.64 ± 0.005
A-3	*LiDAR*	0.31	180.02 ± 0.001	0.74 ± 0.001	-89.68 ± 0.003
	Average-Height	0.36	180.01 ± 0.001	0.62 ± 0.001	-89.72 ± 0.004
B-1	*LiDAR*	0.41	180.17 ± 0.002	0.69 ± 0.001	-89.58 ± 0.003
	Average-Height	0.49	180.14 ± 0.002	0.62 ± 0.002	-89.60 ± 0.005
B-2	*LiDAR*	0.42	179.21 ± 0.002	0.35 ± 0.001	-90.01 ± 0.003
	Average-Height	0.47	179.20 ± 0.002	0.24 ± 0.002	-90.01 ± 0.005

671 The other four evaluation measures for comparing *LiDAR-SfM* and *Average-Height-*
672 *SfM* are presented in Table 5. It can be seen from the table that the RMSE of the back-
673 projection errors using both SfM strategies are at same level (around 2 pixels, i.e., 1.2 cm
674 on the ground). This is expected as the conducted iterative BA-based outlier removal
675 step stops when the RMSE of back-projection errors is smaller than 2.5 pixels. With the
676 same level of back-projection errors, the performance of the two strategies can be
677 compared using the number of images with insufficient tracked features, object points,
678 and tracked SIFT features. As shown in the Table 5, in average, results from *LiDAR-SfM*
679 have a larger number of images with more than 20 SIFT tie points. Moreover, number of
680 object points and SIFT features resulting from the *LiDAR-SfM* is 3-4 times more than
681 that from the *Average-Height-SfM*. These improvements verify that using relatively

682 accurate height information from LiDAR data can help derive reliable search window in
683 the feature matching stage. As a result, more correct matches are established. Figure 13
684 shows an example of derived image-based point clouds for the A-1 dataset using the
685 two strategies. By looking at Figure 13, one can observe that the proposed SfM
686 framework was able to reconstruct more points on the roof that has the highest
687 elevation in the study site (red points in Figure 13a), when compared to the *Average-*
688 *Height-SfM* (Figure 13b). Moreover, it can be observed that the *LiDAR-SfM* outperforms
689 its counterpart in terms of reconstructed points at the periphery of the scene. This is
690 attributed to the fact that the impact of inaccurate height information on the image
691 matching search window is more severe at the edge areas due to the large incident
692 angle (i.e., angle between the light ray and normal vector of the image plane) at those
693 locations.

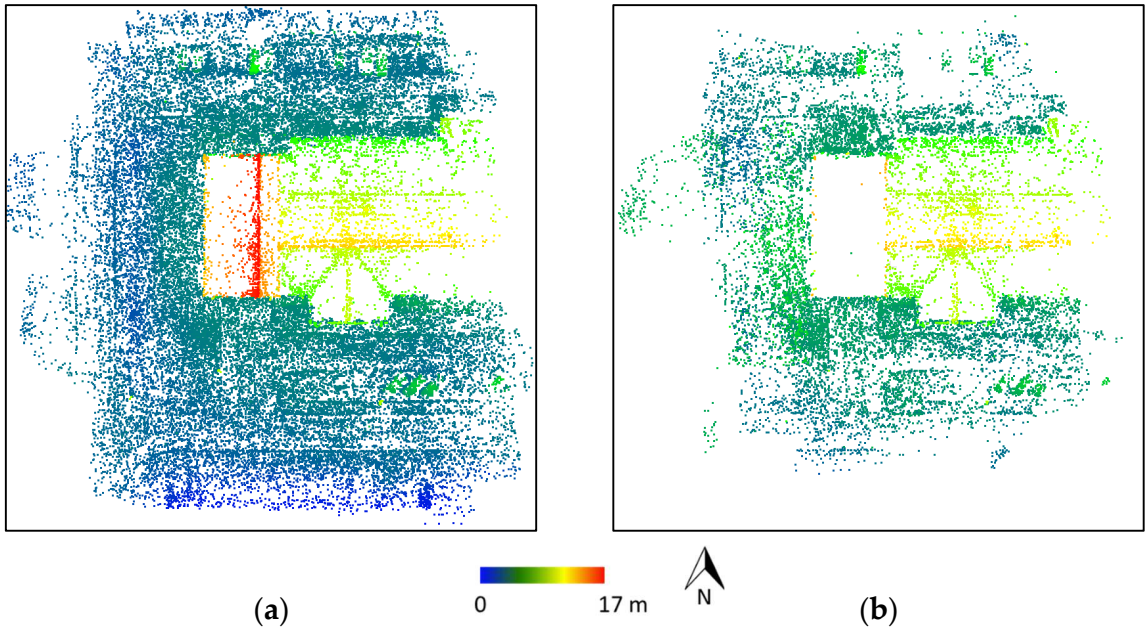


Figure 13. Reconstructed object points (colored by height) of the study site from (a) the proposed *LiDAR-SfM* and (b) *Average-Height-SfM* for the A-1 dataset.

694 Table 5. Performance comparison of the *LiDAR-SfM* and *Average-Height-SfM* in terms of the back-projection errors,
695 number of images with insufficient tracker features, number of reconstructed object points and tracked SIFT features, as
696 well as processing time for the five datasets.

Dataset	SfM Strategy	RMSE of back-projection errors (pixel)	Number of images with insufficient tracked SIFT features	Number of object points (K)	Number of tracked SIFT features (K)	Processing Time (min)					
						LiDAR Generation	Feature Extraction	Ray Tracing	Feature Matching	BA-based Outlier Removal	Total
A-1	*LiDAR*	2.1	2	61.5	230.7	3.3	3.4	4.2	10.1	0.3	21.3
	Average-Height	1.9	25	18.7	62.8	N\A[1]		N\A	7.6	0.2	11.2
A-2	*LiDAR*	1.9	5	66.4	261.6	3.4	3.4	2.7	9.2	0.4	19.0
	Average-Height	2.2	25	20.8	70.2	N\A		N\A	6.7	0.2	10.3
A-3	*LiDAR*	1.6	2	92.7	374.6	3.2	3.4	4.6	15.9	0.6	27.6
	Average-Height	1.7	24	33.5	111.8	N\A		N\A	9.9	0.2	13.6
B-1	*LiDAR*	2.1	3	49.5	191.5	3.8	3.6	3.2	14.6	0.4	25.6
	Average-Height	2.1	26	12.9	41.9	N\A		N\A	8.9	0.2	12.7
B-2	*LiDAR*	2.3	3	46.4	186.3	1.8	4.0	1.8	9.7	0.4	17.7
	Average-Height	2.4	18	12.9	42.9	N\A		N\A	8.0	0.2	12.2

697 **[1] Not applicable.**

As for the processing time, given that the execution time for feature tracking, RANSAC-based and LiDAR-based outlier removal steps is negligible (within 10 seconds), they are not listed in Table 5. Overall, the *LiDAR-SfM* runs slower than the *Average-Height-SfM*. This is mainly due to the fact that down-sampled LiDAR point cloud reconstruction and deriving height information through ray tracing are included in the proposed strategy. In addition, feature matching and iterative BA steps take longer time in the proposed strategy as more conjugate features are derived using LiDAR data. In general, total processing time of *LiDAR-SfM* is approximately twice that of the *Average-Height-SfM*. However, the proposed strategy reconstructs 3-4 times the number of object points and SIFT features than the one using average height. Using the A1 dataset as an example, the total processing time is 0.35 and 0.60 min/thousand object points for the *LiDAR-SfM* and *Average-Height-SfM*, respectively. Therefore, considering the number of reconstructed object points and established conjugate features, the proposed method can be considered to be more efficient.

In general, using LiDAR data in the GNSS/INS-assisted SfM improves the performance over areas with large height variations. The proposed method can efficiently generate significant number of object points and conjugate features with high quality. Moreover, the *LiDAR-SfM* derived point cloud has more coverage on the ground when compared to the *Average-Height-SfM*. With well-distributed object points covering a larger area of the study site, more LiDAR planar patches with different

718 aspects from various flight lines can be extracted. Thus, the performance of subsequent
719 camera/LiDAR integration can be improved.

720 **4.2. Experimental Results for Camera/LiDAR Integration**

721 The proposed tightly-coupled camera/LiDAR integration is conducted on the five
722 datasets. As mentioned in Section 3, the integration process is conducted individually
723 for the A-1, A-2, and A-3 datasets, which were collected on different dates using the
724 same system. In addition, to validate the capability of the proposed strategy in
725 integrating camera and LiDAR data from multiple systems, the B-1 and B-2 datasets –
726 which were collected using two different systems on the same date – are processed
727 simultaneously.

728 The selected threshold values for the automated identification of camera/LiDAR
729 feature correspondences are presented in Table 6. As mentioned in Section 2.3, the
730 proposed camera/LiDAR integration workflow are conducted iteratively. Different
731 thresholds should be used in different iterations to ensure both the computational
732 efficiency and integration accuracy. In the first iteration of integrating process, fewer
733 number of points from both image-based and LiDAR point clouds are used for
734 identifying correspondences. Due to potential poor alignment caused by inaccurate
735 trajectory/system calibration parameters among image-based object points and LiDAR
736 point clouds from different flight lines, more relaxed maximum distance is selected for
737 the first iteration. Also, the search radius is set to a relatively large value to ensure

738 LiDAR points from a given flight line can be included in the spherical region. As the
739 alignment improves throughout the iterations, a greater number of image-based and
740 LiDAR 3D points and more strict thresholds for maximum distance and search radius
741 are used to achieve the best integration results. Figure 14 shows an example of
742 established primitives in iterations 1, 2, and 3 for the A-1 dataset. It can be seen from
743 Figure 14a to Figure 14c that the size of the planar patches decreases throughout the
744 iteration while the number of established primitives increases. Looking into a sample
745 primitive from these iterations in Figure 14d to Figure 14f, the discrepancy among
746 image-based point and LiDAR points from different flight lines is quite large in the first
747 iteration, and it significantly improves after two iterations.

748 Table 6. Selected threshold values throughout the different iterations for the proposed
749 automated identification of camera/LiDAR feature correspondences.

Threshold Description		**Number of Iteration**		
		1	2	≥ 3
Distance threshold for image-based point cloud down-sampling		2.0 m	1.5 m	1.0 m
LiDAR point cloud down-sampling	*VLP-system*	8 out of 32 channels; every 2nd firings	16 out of 32 channels; every 2nd firings	All channels; every 2nd firings
	Riegl-System	every 10th firings	every 5th firings	every 2nd firings
Maximum distance for an image-based 3D point to its closest LiDAR point for deriving planar patch		3.0 m	1.0 m	0.3 m
Search radius for extracting neighboring LiDAR points		1.0 m	0.5 m	0.3 m
Criteria for a derived planar patch to be	Minimum number of inliers	30	50	50
	Minimum ratio of inliers	50 %		

considered as valid	Maximum RMSE of normal distances	30 cm

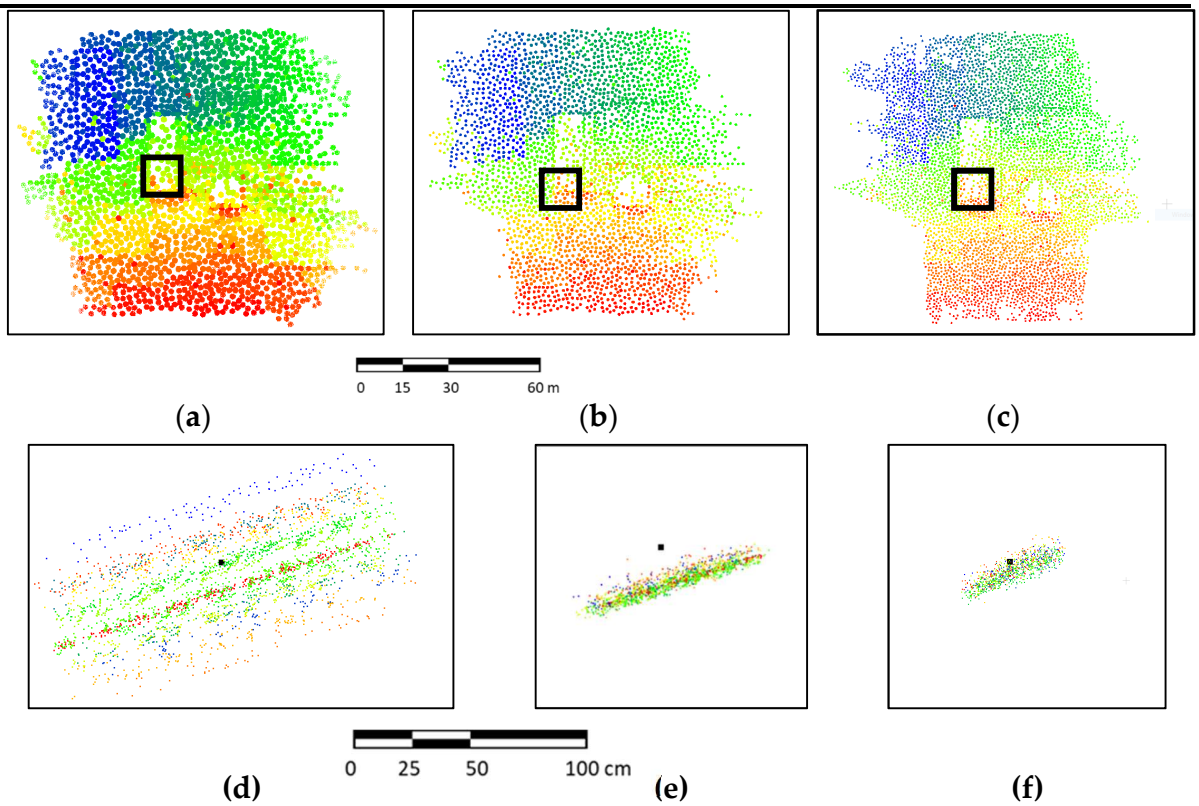


Figure 14. An example of established primitives for the A-1 dataset: top view of established primitives in (a) iteration 1 with 2,090 primitives, (b) iteration 2 with 2,891 primitives, and (c) iteration 3 with 4,706 primitives, where the LiDAR points are colored according to the primitive index; side-view of a sample primitive (shown in black box) from (d) iteration 1, (e) iteration 2, and (f) iteration 3, where the black point denotes the image-based 3D point and other LiDAR points are colored according to the flight line they originate from.

In the integrated-BA procedure, second order of polynomial is selected to represent the position and orientation parameters while three neighboring reference points are used to interpolate the trajectory information for a given image/LiDAR observation (i.e., $p = 2, q = 3$). As for a-priori variances for the observations, the values for image coordinate measurements and GNSS/INS information are identical to those utilized in

the GNSS/INS-assisted BA stage of the *LiDAR-SfM* (Section 4.1) – i.e., (7 pixels)^2 for the SIFT features; (3 cm)^2, (0.025°)^2, and (0.08°)^2, for the position, roll/pitch angles, and heading angle, respectively. The a-priori variance for LiDAR point is set to (6 cm)^2 based on their expected accuracy. In this study, the following system calibration parameters are estimated in the integrated-BA: i) camera IOP including principal distance (c) and distortion parameters (K_1, K_2, P_1, P_2), ii) camera boresight angles ($\Delta\omega_c, \Delta\phi_c, \Delta\kappa_c$), and iii) LiDAR boresight angles ($\Delta\omega_{lu}, \Delta\phi_{lu}, \Delta\kappa_{lu}$) as well as lever arm components in X and Y directions ($\Delta X_{lu}, \Delta Y_{lu}$). Previously conducted bias impact analysis (Kersting et al., 2011) shows that in a camera system, there is a high correlation between the planimetric lever arm components and boresight roll and pitch angles. Also, the principal distance (c) of the camera is highly correlated with the lever arm Z component. Hence, the camera lever arm components are excluded from the unknown parameters in the BA process. Moreover, the Z lever arm component for LiDAR is fixed as it requires vertical control (Ravi et al., 2018b). However, in the case of simultaneous integration of multiple systems, by fixing ΔZ_{lu} for a LiDAR sensor in a given system, the Z lever arm component for laser units onboard other UAVs can be estimated. Now that the involved thresholds for the iterative process and system calibration parameters that will be estimated in the integrated-BA have been explained, experimental results for the camera/LiDAR integration and accuracy analysis for the five datasets will be presented in the remainder of this subsection.

775 The estimated system calibration parameters along with their standard deviations
776 resulting from the proposed camera/LiDAR integration process are presented in Table 7.
777 In this table, the derived square root of a-posteriori variance factor ($\hat{\sigma}_0$), RMSE value of
778 the back-projection errors, and RMSE of the distances from LiDAR points to their
779 corresponding adjusted planar patch are also reported. It is worth mentioning that even
780 though the integration process is simultaneously conducted for the B-1 and B-2 datasets,
781 the RMSE values are derived separately for each system. In other words, image tie
782 features and LiDAR points belonging to a given dataset are used to compute the back-
783 projection errors and LiDAR point to plane distances, respectively. As presented in
784 Table 7, the $\hat{\sigma}_0$ values for the four experiments are smaller than 1, which reveals that the
785 assigned a-priori variances for the image/LiDAR observations and trajectory
786 information are conservative. This is also verified by the reported RMSE values of back-
787 projection errors, which are around 2-3 pixels (i.e., approximately 1.5 cm on the ground)
788 – more accurate than the expected accuracy of the SIFT features of 7 pixels – for all
789 datasets. As for the LiDAR sensors, RMSE of LiDAR point to plane distances serves as
790 an indication of the alignment among LiDAR points from multiple flight lines. The
791 RMSE values for the datasets collected using *VLP-system* are within the range accuracy
792 of the LiDAR unit (±3 cm). For the B-2 dataset captured by *Riegl-system*, the RMSE of
793 LiDAR point to plane distances is 1.33 cm, which is better than the other four datasets.
794 This is because of the lower noise level of the Riegl miniVUX-1DL laser scanner

795 compared to the Velodyne VLP-32C, with the former having a ±1.5 cm range accuracy.
796 Given that the RMSE values for both LiDAR sensors are smaller than their range
797 accuracy, LiDAR points from multiple flight lines are expected to be well-aligned after
798 the integration. It can be seen from Table 7 that standard deviation values for the
799 estimated system calibration parameters – including IOP and camera/LiDAR mounting
800 parameters – are small. Moreover, it has been observed that the estimated system
801 calibration parameters are not highly correlated except for the radial distortion
802 coefficients K1 and K2, due to their similar impact on 2D feature coordinates in the
803 image plane. Based on the low correlation and small standard deviations of the
804 estimated system calibration parameters as well as the RMSE values for image/LiDAR-
805 based residuals shown in Table 7, one can conclude that the system calibration
806 parameters are accurately estimated using the proposed integration strategy.

807 As mentioned earlier, GNSS/INS trajectory is refined in the integration process using
808 the concept of reference points. Table 8 presents the mean, standard deviation, and
809 RMSE of the differences between initial and refined position and orientation parameters
810 for the valid reference points – i.e., the ones that have corresponding image/LiDAR
811 observations. The mean difference values for all parameters are close to zero for all
812 datasets. The RMSE values for the position differences are smaller than 3 cm, which are
813 at the same level of the nominal positional accuracy of the APX-15. The same

814 performance can be observed for the orientation parameters – the RMSE values for
815 roll/pitch angles are in the range of

816 Table 7. Original and estimated system calibration parameters using the proposed
817 camera/LiDAR integration strategy for the five datasets, as well as the RMSE values of
818 the back-projection errors and LiDAR point to plane distances.

Datasets		**A-1**	**A-2**	**A-3**	**B-1 & B-2**	
Systems		*VLP*	*VLP*	*VLP*	*VLP*	*Riegl*
$\widehat{\boldsymbol{\sigma}}_{\mathbf{0}}$		0.42	0.41	0.43	0.35	
RMSE of back-projection errors (pixels)		2.78	2.51	2.35	3.09	2.93
RMSE of LiDAR point to plane distances (cm)		2.47	2.43	2.53	2.50	1.33
Camera	c (pixel)	8023.18 ± 0.83	8026.07 ± 0.82	8023.93 ± 0.83	8013.21 ± 0.83	7532.60 ± 0.84
	$K1$ (10^{-10} pixel^{-2})	8.53 ± 0.18	8.46 ± 0.18	8.41 ± 0.18	8.16 ± 0.18	-12.09 ± 0.21
	$K2$ (10^{-17} pixel^{-4})	-5.46 ± 0.08	-5.45 ± 0.08	-5.44 ± 0.08	-5.52 ± 0.08	-2.51 ± 0.10
	$P1$ (10^{-7} pixel^{-1})	0.97 ± 0.04	1.11 ± 0.04	1.04 ± 0.04	1.24 ± 0.04	-1.22 ± 0.05
	$P2$ (10^{-8} pixel^{-1})	-9.66 ± 0.63	-9.90 ± 0.60	-9.32 ± 0.61	-10.98 ± 0.59	-0.68 ± 0.64
	$\Delta\omega_c$ (°)	180.44 ± 0.01	180.05 ± 0.01	180.03 ± 0.01	180.21 ± 0.01	179.14 ± 0.01
	$\Delta\phi_c$ (°)	0.77 ± 0.01	0.76 ± 0.01	0.76 ± 0.01	0.80 ± 0.01	0.36 ± 0.01
	$\Delta\kappa_c$ (°)	-89.55 ± 0.01	-89.57 ± 0.01	-89.65 ± 0.01	-89.58 ± 0.01	-90.01 ± 0.01
LiDAR	ΔX_{lu} (cm)	-10.45 ± 0.25	-11.52 ± 0.23	-10.72 ± 0.25	-11.09 ± 0.16	20.16 ± 0.13
	ΔY_{lu} (cm)	3.60 ± 0.19	3.49 ± 0.16	3.30 ± 0.19	3.22 ± 0.12	-1.35 ± 0.16
	ΔZ_{lu} (cm)	N\A	N\A	N\A	N\A	2.76 ± 0.03
	$\Delta\omega_{lu}$ (°)	0.56 ± 0.01	0.58 ± 0.01	0.53 ± 0.01	0.54 ± 0.01	-90.13 ± 0.01
	$\Delta\phi_{lu}$ (°)	-0.22 ± 0.01	-0.21 ± 0.01	-0.21 ± 0.01	-0.23 ± 0.01	0.18 ± 0.01
	$\Delta\kappa_{lu}$ (°)	-0.21 ± 0.01	-0.22 ± 0.01	-0.24 ± 0.01	-0.12 ± 0.01	-90.14 ± 0.01

819

820 0.01° to 0.02° while those for heading angle is within 0.08°. In summary, since the
821 differences between the refined and original trajectory are compatible with the expected
822 accuracy, it can be concluded that: i) the accuracy of the GNSS/INS unit (APX) is in
823 agreement with the sensor specifications and ii) the selected polynomial order (i.e.,
824 second order) and number of neighboring reference points (i.e., three) for trajectory
825 interpolation are reasonable.

826 Table 8. Mean, STD, and RMSE of the differences between initial and refined trajectory
827 information for the five datasets.

Dataset	Number of Valid Reference Points	Statistics Criteria	X_{dif} (cm)	Y_{dif} (cm)	Z_{dif} (cm)	ω_{dif} (°)	ϕ_{dif} (°)	κ_{dif} (°)
A-1	5,206	Mean	0.00	0.00	0.00	0.00	0.00	0.00
		STD	1.41	1.50	1.37	0.01	0.01	0.07
		RMSE	1.41	1.50	1.37	0.01	0.01	0.07
A-2	5,152	Mean	0.00	0.00	0.00	0.00	0.00	0.00
		STD	1.42	1.49	1.28	0.01	0.01	0.07
		RMSE	1.42	1.49	1.28	0.01	0.01	0.07
A-3	5,152	Mean	0.00	0.00	0.00	0.00	0.00	0.00
		STD	1.81	1.43	1.44	0.01	0.01	0.07
		RMSE	1.81	1.43	1.44	0.01	0.01	0.07
B-1	5,161	Mean	0.09	-0.11	0.03	0.00	0.00	0.00
		STD	1.61	1.77	2.52	0.02	0.01	0.08
		RMSE	1.61	1.78	2.53	0.02	0.01	0.08
B-2	5,369	Mean	-0.08	0.11	-0.04	0.00	0.00	0.00
		STD	1.23	1.01	1.26	0.01	0.01	0.04
		RMSE	1.23	1.02	1.26	0.01	0.01	0.04

828 In the remainder of this section, the relative and absolute accuracy of the generated
829 image/LiDAR data after integrated-BA are qualitatively and quantitatively analyzed. To
830 do so, three additional products are generated:

- Image-based dense point cloud: Although image-based sparse point cloud is available from the integrated-BA, a well-distributed dense point cloud is generated for conducting a more comprehensive and reliable comparison with the LiDAR point cloud. The Patch-based multi-view stereo (PMVS) algorithm proposed by Furukawa and Ponce (2009) is conducted using estimated system calibration parameters and refined trajectory for dense point cloud generation,
- Image-based coordinates for CPs: The centers of the twelve checkerboard targets are manually identified in the visible images. Then, using the estimated camera IOP, mounting parameters, and refined trajectory information, 3D coordinates of the target centers (i.e., CPs) are estimated through a multi-light ray intersection,
- LiDAR-based coordinates for CPs: Centers of the highly reflective checkerboard targets are first manually identified from the generated LiDAR point cloud based on the intensity information, and denoted as initial points. The initial points are expected to have an accuracy of ±3 to ±5 cm in the horizontal directions due to the noise level of the LiDAR data. Then, the strategy proposed in Section 2.2 is used to derive the best-fitting plane in the neighborhood of the initial points, and reliable Z coordinates are derived by projecting the initial points onto the defined planes.

The image-based dense point cloud will be qualitatively and quantitatively compared with the LiDAR point cloud to evaluate the relative accuracy of the integrated-BA results. In addition, through a comparison between image/LiDAR-based

coordinates for CPs with the RTK-GNSS values, absolute accuracy of the image-based and LiDAR point clouds can be verified.

4.2.1. Qualitative Analysis of the Integrated-BA Results

The integrated-BA results from the B-1 and B-2 datasets are used to conduct the qualitative analysis since alignment among image-based/LiDAR point clouds from multiple systems can be evaluated simultaneously. The relative accuracy of the integrated-BA results is qualitatively analyzed through profiles extracted from both image-based dense point clouds and LiDAR point clouds for both *VLP-system* and *Riegl-system*. Figure 15 shows the locations as well as the side-view of four extracted profiles (3 m by 0.2 m). It is clear from the profiles that the noise level from Riegl laser unit is lower than that from VLP32C. Compared with the LiDAR points, image points from dense point clouds are much sparser with more noise. This is more obvious in the image object points from the roof area (Profiles 3 and 4). Although with different noise level, overall, the image-based object points and LiDAR points from the two systems are well-aligned – i.e., no shift exists among these point clouds.

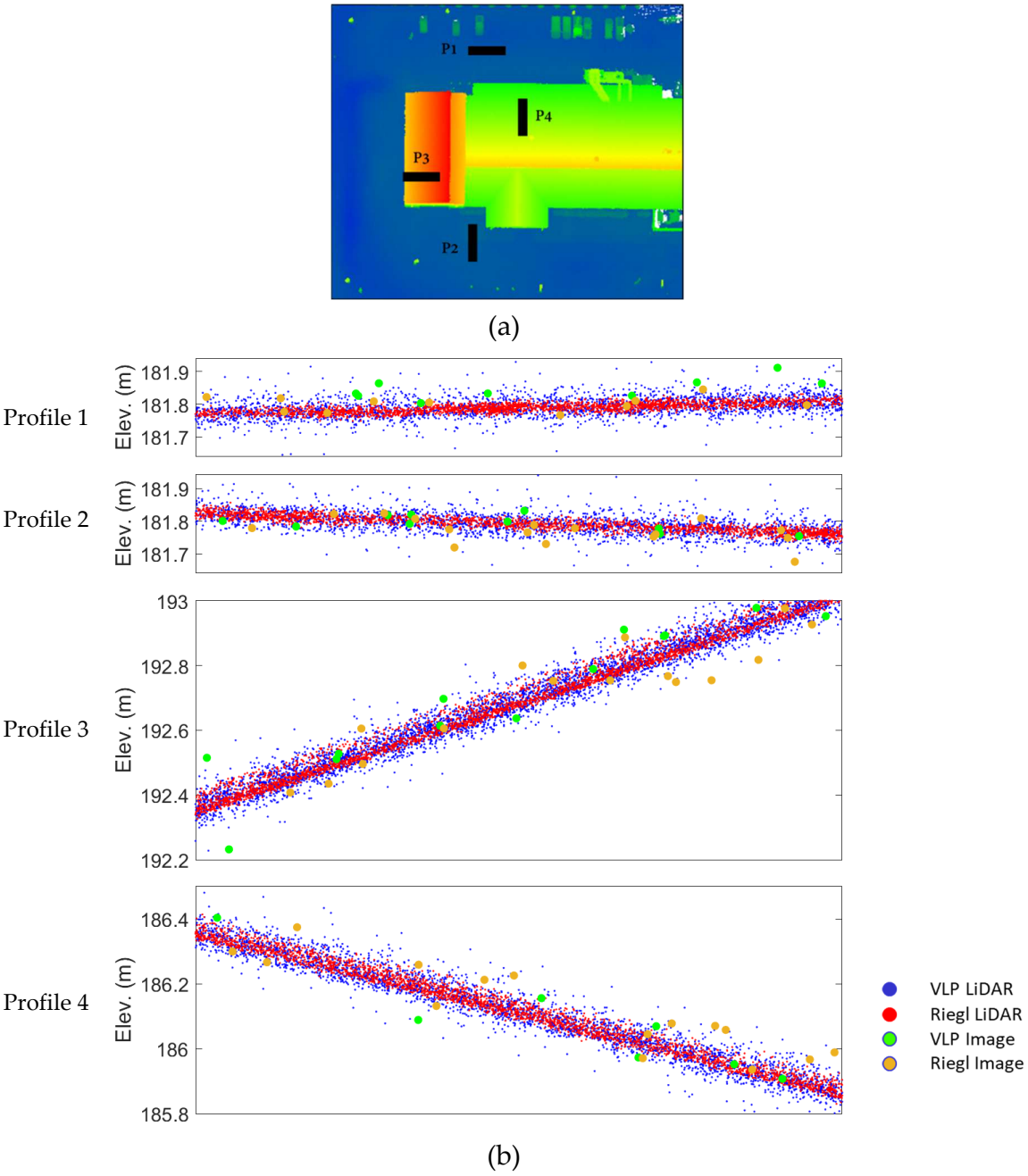
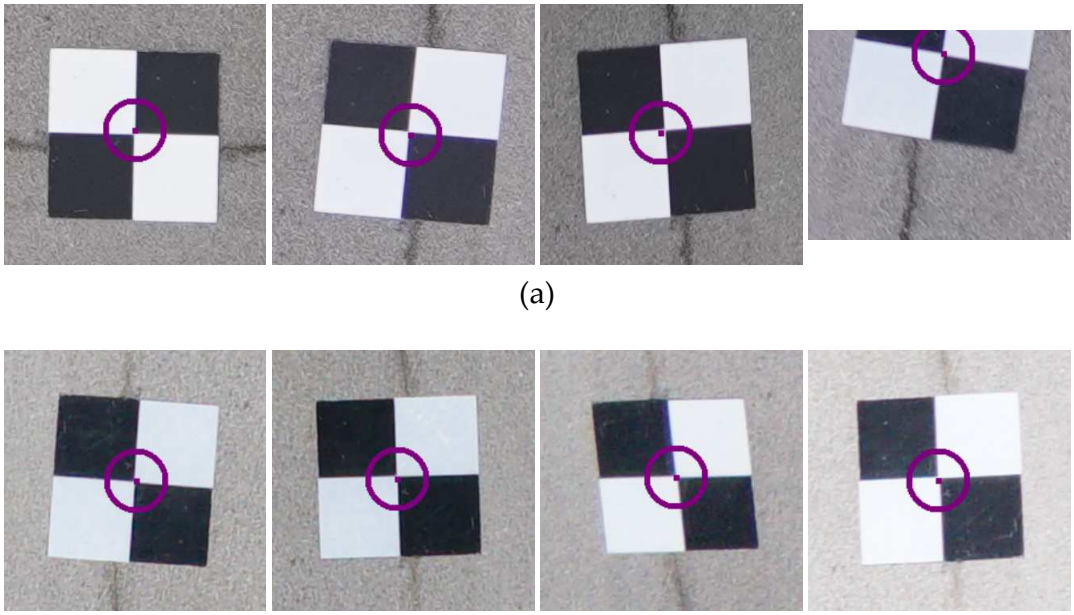


Figure 15. Relative accuracy analysis for the B-1 and B-2 datasets: (a) locations of the four extracted profiles and (b) side-view of those profiles.

Established checkerboard targets are used as CPs to qualitatively evaluate the absolute accuracy of the integrated-BA results. With the center of the targets surveyed using RTK-GNSS, a 2D analysis is conducted by back-projecting the target center to the

image space using the refined EOP and IOP. The compatibility between the back-projected point and the target center in image serves as an indication of the quality of the estimated system calibration parameters and refined trajectory. Figure 16 shows an example of the 2D analysis, where the center of a target is back-projected to some images from both *VLP-system* and *Riegl-system*. This figure presents back-projection results on four randomly selected visible images. As shown in the figure, back-projected target centers (represented by magenta markers) and the actual centers in the captured images are in agreement within 3 pixels for both systems. This suggests that the estimated IOP and refined EOP are accurate.



(a)

(b)

Figure 16. 2D absolute accuracy analysis for image-based point cloud for the (a) B-1 and (b) B-2 datasets, where back-projected points of target centers on the images are represented by magenta markers.

878 The absolute accuracy of the experimental results can also be evaluated in 3D space
879 by checking the alignment among RTK-GNSS and image-based ground coordinates of
880 the target center and LiDAR points representing the same target. Figure 17 shows such
881 alignment in the horizontal and vertical directions. Figures 17a and 17b show top view
882 of LiDAR point clouds (colored by intensity value) from *VLP-system* and *Riegl-system*
883 along with the image-based and RTK-GNSS-derived centers of the target (shown by
884 orange/green points). It can be observed that these two points are very close to each
885 other in horizontal directions. Moreover, the expected centers defined by the LiDAR
886 points are also in agreement with the RTK-GNSS ones. The vertical accuracy of the
887 image-based and LiDAR point clouds is verified by Figure 17c. In this figure, the Z
888 coordinates of the image-based target centers from *VLP-system* and *Riegl-system* (colored
889 in yellow and magenta) are compatible with the RTK-GNSS in the range of 2 cm. The
890 figure also shows that LiDAR points from the two systems are well-aligned with the
891 RTK-GNSS-based target centers. However, different from what was observed in Figure
892 15, the noise level of LiDAR points from Riegl laser unit is relatively high. Such
893 behavior is hypnotized to be caused by the nature of LiDAR pulse returns from highly
894 reflective surfaces. Figure 17d shows a sample of Riegl-derived LiDAR point cloud at
895 the transition between the target and ground area. It can be seen from this figure that
896 LiDAR points from ground area are much cleaner compared to those on the target. This

observation verifies the hypothesis that noisy LiDAR points over the target is attributed to the nature of LiDAR

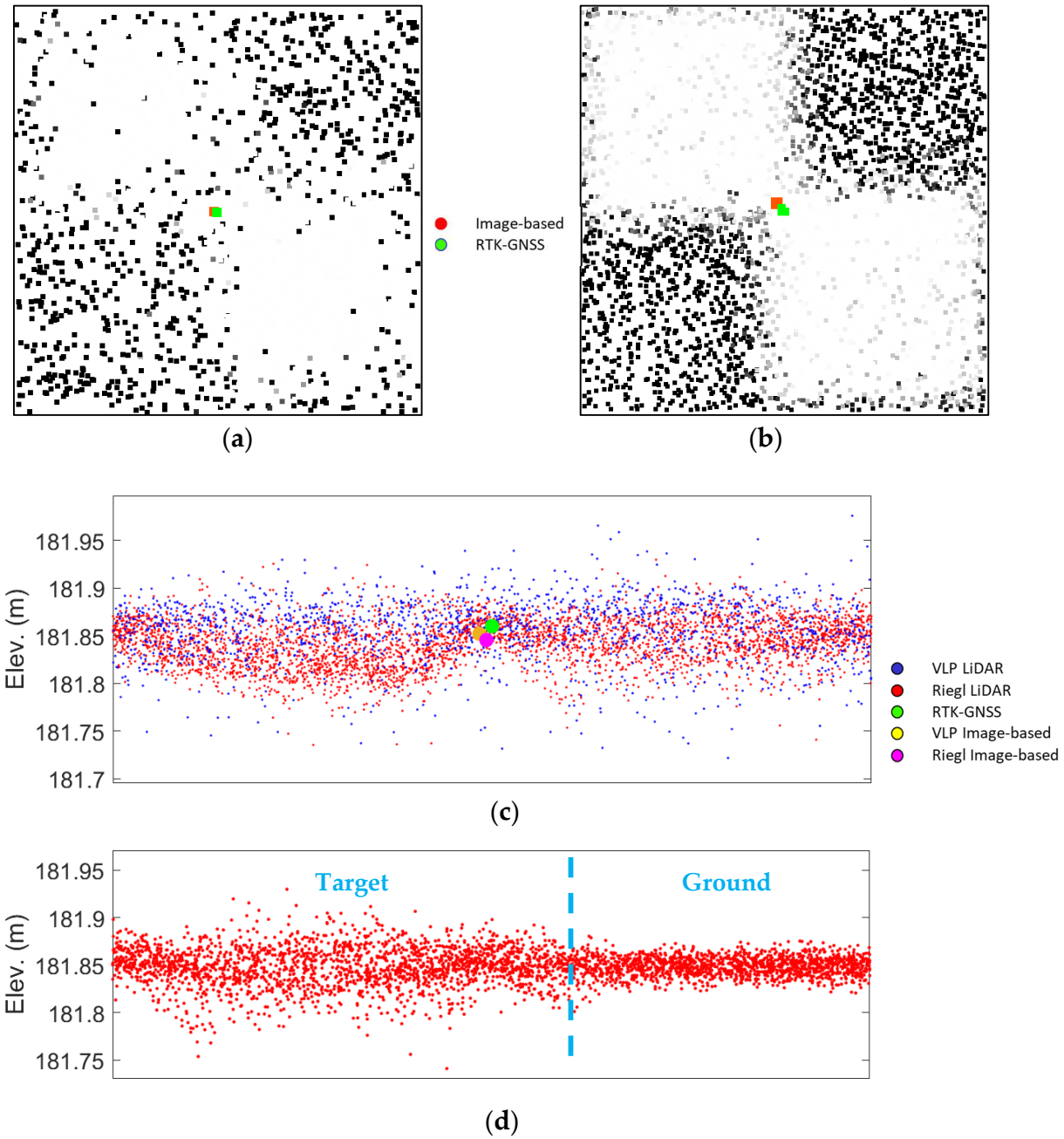


Figure 17. 3D aboslute accuracy analysis for the B-1 and B-2 datasets: (a) and (b) top view of RTK-GNSS/image-based target centers and LiDAR points for *VLP-system* and *Riegl-system*, respectively; (c) side-view of RTK-GNSS/image-based target centers and

LiDAR points for both systems; and (d) side-view of a sample Riegl-derived LiDAR point cloud at the transition between the target and ground area (points on the left side and right side of the blue dash line belong to the target and ground area, respectively).

pulse interaction with the reflective surface. In summary, the qualitative analysis shows that image and LiDAR data from the B-1 and B-2 datasets after the integrated-BA process has reasonable absolute/relative accuracy.

4.2.2. Quantitative Analysis of the Integrated-BA Results

For the quantitative analysis, relative accuracy between the image-based and LiDAR point clouds is first evaluated. Here, sparse and dense image-based point clouds are both compared to LiDAR point clouds. The discrepancy between the image-based and LiDAR point clouds is derived through the LSA model introduced in our previous work (Zhou et al., 2020). In brief, correspondences between the two point clouds are first established using the same approach introduced in Section 2.2. A modified weight matrix-based LSA is then used to estimate the net X, Y, and Z discrepancies (dx, dy, dz). The estimated discrepancies and corresponding $\hat{\sigma}_0$ for the five datasets are presented in Table 9. The derived $\hat{\sigma}_0$ in the LSA procedures represents the average absolute distance between the image-LiDAR point pairs as an identity weight matrix is adopted for the observations. As reported in Table 9, the average absolute distance is in the range of 6-7 cm, and 4-6 cm for image-based sparse and dense point clouds, respectively. This suggests that the overall noise level of the sparse point cloud is higher than that in the

dense point cloud. The estimated vertical discrepancy for the five datasets is within 1 cm, while the horizontal discrepancies are larger. For the sparse point cloud, the horizontal discrepancies are in the range of -2 cm to 2 cm, which are smaller than that from dense point cloud (i.e., -3 cm to 5 cm). A PCA-based dimensionality analysis is carried out on the dispersion matrix of the estimated net discrepancy vector. The principal components represent directions of the discrepancies that exhibit a maximum amount of variance in the mapping frame, while the corresponding eigenvalues denote the variances along these directions (Zhou et al., 2020). The analysis results – which are similar for all datasets – show that the Z discrepancy estimate is the most reliable component since the corresponding variance only accounts for less than 1% of the total variance. This statement is also verified by the standard deviations of the discrepancies – the standard deviation values for the estimated X and Y discrepancies are 5-10 times larger than that for Z discrepancy. These variations in the variances are attributed to the fact that the majority of reconstruction areas are on horizontal surfaces. Point pairs belonging to horizontal areas will contribute to the estimation of the Z discrepancy. The horizontal discrepancy is mainly derived by points reconstructed on the building roof area. Since dense point clouds include more object points on the roof compared to sparse point clouds, the former have smaller standard deviations for horizontal discrepancies. However, it should be noted that the noise level of the generated image-based point cloud over the roof area is higher than other parts of the study site due to

the homogeneous texture of the gable roof. This statement has also been verified by comparing the profiles from the ground (Profiles 1 and 2) and roof areas (Profiles 3 and 4) as shown in Figure 15. Having more object points on the roof area with larger noise level, the X and Y discrepancies derived from dense point cloud are larger compared to the ones from sparse point cloud. In conclusion, the comparison between image-based and LiDAR point clouds verifies the relative vertical accuracy and horizontal accuracy – to a lesser degree – of the integrated-BA.

Table 9. Estimated X, Y, and Z discrepancies between the image-based sparse/dense point clouds and LiDAR point clouds for the five datasets after camera/LiDAR integrated-BA.

Dataset	Image-based sparse vs. LiDAR				Image-based dense vs. LiDAR			
	$\hat{\sigma}_0$ (cm)	dx (cm)	dy (cm)	dz (cm)	$\hat{\sigma}_0$ (cm)	dx (cm)	dy (cm)	dz (cm)
A-1	6.77	-2.07 ± 0.24	2.39 ± 0.20	0.22 ± 0.02	5.19	-3.32 ± 0.13	5.26 ± 0.14	0.77 ± 0.02
A-2	6.07	0.25 ± 0.23	1.22 ± 0.22	0.04 ± 0.02	4.46	-2.69 ± 0.11	2.56 ± 0.12	0.47 ± 0.02
A-3	6.12	-0.90 ± 0.23	-0.70 ± 0.21	0.08 ± 0.02	5.46	-1.27 ± 0.14	2.88 ± 0.14	0.89 ± 0.02
B-1	7.62	-0.86 ± 0.26	0.89 ± 0.25	0.06 ± 0.03	6.24	-2.64 ± 0.18	-2.44 ± 0.17	1.04 ± 0.02
B-2	7.56	0.36 ± 0.25	1.19 ± 0.22	-0.62 ± 0.03	5.32	-1.91 ± 0.14	1.91 ± 0.14	0.53 ± 0.02

Absolute accuracy of the image/LiDAR products is evaluated using the CPs. Table 10 reports the mean, STD, and RMSE differences between the image-based/LiDAR-based and RTK-GNSS coordinates of the twelve CPs for the five datasets. According to

949 the statistics reported in Table 10, an absolute accuracy of 1-2 cm has been achieved for
950 image-based point cloud in all directions from the five datasets. It is also clear from
951 Table 10 that the vertical accuracy of LiDAR point cloud is in the range of 1-3 cm.
952 However, the absolute differences in the horizontal directions range from 2 cm to 4 cm
953 for LiDAR point cloud, which is worse compared to the vertical accuracy. The larger
954 differences in the horizontal direction are mainly due to the difficulty in identifying the
955 center of the targets in LiDAR point cloud based on the intensity information. Also, the
956 horizontal accuracy of LiDAR is known to be worse than the vertical accuracy. In
957 summary, both image-based and LiDAR point clouds achieve high absolute accuracy
958 after the integrated-BA.

959 Table 10. Mean, STD, and RMSE of the differences between image-based/LiDAR-based
960 and RTK-GNSS coordinates of the twelve CPs for the five datasets after camera/LiDAR
961 integrated-BA.

Dataset	**Statistics Criteria**	**Image-based vs. RTK-GNSS**			**LiDAR-based vs. RTK-GNSS**		
		X_{dif} (cm)	Y_{dif} (cm)	Z_{dif} (cm)	X_{dif} (cm)	Y_{dif} (cm)	Z_{dif} (cm)
A-1	Mean	-0.2	0.1	0.0	-3.5	0.5	0.0
	STD	1.3	0.8	1.9	2.3	3.3	1.6
	RMSE	1.3	0.8	1.9	4.2	3.3	1.5
A-2	Mean	-2.4	0.6	0.4	-0.7	1.9	-0.4
	STD	0.6	1.2	1.4	2.5	2.7	1.4
	RMSE	2.4	1.3	1.4	2.6	3.3	1.4
A-3	Mean	-2.0	0.0	-1.7	-1.7	2.6	-3.1
	STD	0.8	1.4	1.8	3.1	3.4	1.7
	RMSE	2.1	1.3	2.4	3.5	4.3	3.5
B-1	Mean	-2.2	-0.2	-1.0	1.3	-2.1	-0.6
	STD	1.0	1.5	1.4	2.4	2.2	1.7

	RMSE	2.4	1.4	1.6	2.7	3.0	1.8
B-2	Mean	-0.9	-0.8	-1.3	2.3	3.1	-1.3
	STD	1.0	1.0	1.2	1.9	2.9	1.5
	RMSE	1.3	1.3	1.7	3.0	4.2	2.0

5. Conclusions and Recommendations for Future Work

In this paper, a fully automated, tightly-coupled camera/LiDAR integration workflow for GNSS/INS-assisted UAV systems has been proposed. Rather than relying on targets or manually identifying conjugate features in camera and LiDAR data, correspondences between these two data sources are automatically extracted. More specifically, a LiDAR/GNSS/INS-assisted SfM is first developed for image-based point cloud generation. The proposed SfM strategy overcomes the limitations of the state-of-the-art SfM algorithms by exploiting available trajectory information as well as LiDAR data to identify conjugate features among overlapping images. The image-based object points are then used as seed points to identify corresponding planar patches in different LiDAR tracks. This automated correspondence identification strategy is more efficient compared to traditional point-to-point algorithms. With the established correspondences, an integrated-BA procedure that simultaneously estimates the involved system calibration parameters for camera and LiDAR sensors while refining the trajectory is conducted. The developed BA model can also be generalized to support various platforms and sensors. The integration process is conducted iteratively where different thresholds are used within the different iterations to ensure both the

computational efficiency and integration accuracy. Five experimental datasets were used to evaluate the performance of the developed integration strategy. Below are the key findings of the experimental results from the presented strategy:

- LiDAR/GNSS/INS-assisted SfM strategy outperforms an existing approach utilizing average height in terms of the number of reconstructed object points/tie features, coverage, and distribution of the derived point cloud. This proves that using a relatively accurate 3D model of the object space – a sparse LiDAR point cloud in this study – can significantly improve the performance of the GNSS/INS-assisted SfM for datasets over areas with large height variations.
- The proposed camera/LiDAR integration framework, which eliminates the need for manual intervention or specific targets, can be conducted more frequently and even for individual data collection missions to ensure a good accuracy of the photogrammetric and LiDAR products.
- The integrated-BA results for the five datasets show small standard deviation values for the estimated system calibration parameters as well as low correlation among those parameters. This indicates that the system calibration parameters of both camera and LiDAR sensors are accurately estimated. The estimated system calibration parameters ensure high relative accuracy between the image-based and LiDAR point clouds, particularly in the Z direction. In addition, through a comparison with RTK-GNSS measurements of CPs, absolute accuracy of the image-

based and LiDAR point clouds after integrated-BA is verified, which is in the range of 3 – 5 cm at 41 m flying height. Moreover, the proposed integration process is conducted on two systems simultaneously. Similar accuracy is achieved compared to the results from a single system. This verifies the proposed framework is capable of aligning multiple sensors/systems.

The main limitation of the proposed integration strategy is that GNSS/INS information is loosely-coupled within the process – i.e., GNSS/INS raw measurements are not included in the BA procedure. Instead, trajectory generated from GNSS/INS integration is directly used in the proposed framework. To overcome this issue, future work will focus on tightly-coupled GNSS/INS/camera/LiDAR integration. Future work will also focus on including different types of features, other than planar patches, in the BA to improve the capability of decoupling system calibration parameters and performance of camera/LiDAR integration. Moreover, the integration process will be extended to incorporate other imaging sensors – e.g., hyperspectral push-broom scanners – that can be useful for deriving highly accurate products for various applications such as precision agriculture and forestry.

Acknowledgments

The information, data, or work presented herein was funded in part by the Civil Engineering Center for Applications of UAS for a Sustainable Environment (CE CAUSE) and 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.

References

Applanix. Apx-15 UAV Datasheet. Available online: https://www.applanix.com/downloads/products/specs/APX15_UAV.pdf (accessed on 4 May 2021).

Besl, P. J., & McKay, N. D. (1992, April). Method for registration of 3-D shapes. *In Sensor fusion IV: control paradigms and data structures* (Vol. 1611, pp. 586-606). International Society for Optics and Photonics.

Cai, H., Pang, W., Chen, X., Wang, Y., & Liang, H. (2020). A novel calibration board and experiments for 3D LiDAR and camera calibration. *Sensors*, 20(4), 1130.

Chen, Y., & Medioni, G. (1992). Object modelling by registration of multiple range images. Image and vision computing, 10(3), 145-155.

Elbahnasawy, M. (2018). *GNSS/INS-assisted Multi-camera Mobile Mapping: System Architecture, Modeling, Calibration, and Enhanced Navigation* (Doctoral dissertation, Purdue University).

Fonstad, M. A., Dietrich, J. T., Courville, B. C., Jensen, J. L., & Carbonneau, P. E. (2013). Topographic structure from motion: a new development in photogrammetric measurement. *Earth surface processes and Landforms*, *38(4)*, 421-430.

Furukawa, Y., & Ponce, J. (2009). Accurate, dense, and robust multiview stereopsis. *IEEE transactions on pattern analysis and machine intelligence*, *32*(8), 1362-1376.

Glira, P., Pfeifer, N., & Mandlburger, G. (2016). Rigorous Strip adjustment of UAV-based laserscanning data including time-dependent correction of trajectory errors. *Photogrammetric Engineering & Remote Sensing*, *82*(12), 945-954.

Glira, P., Pfeifer, N., & Mandlburger, G. (2019). HYBRID ORIENTATION OF AIRBORNE LIDAR POINT CLOUDS AND AERIAL IMAGES. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, *4*.

Gneeniss, A. S., Mills, J. P., & Miller, P. E. (2015). In-flight photogrammetric camera calibration and validation via complementary lidar. *ISPRS journal of photogrammetry and remote sensing*, *100*, 3-13.

Guindel, C., Beltrán, J., Martín, D., & García, F. (2017, October). Automatic extrinsic calibration for lidar-stereo vehicle sensor setups. *In 2017 IEEE 20th international conference on intelligent transportation systems (ITSC)* (pp. 1-6). IEEE.

Habib, A., & Morgan, M. (2005). Stability analysis and geometric calibration of off-the-shelf digital cameras. *Photogrammetric Engineering & Remote Sensing*, *71*(6), 733-741.

Hasheminasab, S. M., Zhou, T., & Habib, A. (2020). GNSS/INS-Assisted structure from motion strategies for UAV-Based imagery over mechanized agricultural fields. *Remote Sensing*, 12(3), 351.

Hastedt, H., & Luhmann, T. (2015). Investigations on the quality of the interior orientation and its impact in object space for UAV photogrammetry. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 40.

He, F., & Habib, A. (2016). Automated relative orientation of UAV-based imagery in the presence of prior information for the flight trajectory. Photogrammetric Engineering & Remote Sensing, 82(11), 879-891.

Kersting, A. P., Habib, A., & Bang, K. I. (2011, January). Mounting parameters calibration of GPS/INS-assisted photogrammetric systems. In *2011 International Workshop on Multi-Platform/Multi-Sensor Remote Sensing and Mapping* (pp. 1-6). IEEE.

Li, J., Yang, B., Chen, C., & Habib, A. (2019). NRLI-UAV: Non-rigid registration of sequential raw laser scans and images for low-cost UAV LiDAR point cloud quality improvement. *ISPRS Journal of Photogrammetry and Remote Sensing*, *158*, 123-145.

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

Nex, F., & Remondino, F. (2014). UAV for 3D mapping applications: a review. *Applied geomatics*, 6(1), 1-15.

Pentek, Q., Kennel, P., Allouis, T., Fiorio, C., & Strauss, O. (2020). A flexible targetless LiDAR–GNSS/INS–camera calibration method for UAV platforms. *ISPRS Journal of Photogrammetry and Remote Sensing*, *166*, 294-307.

Pusztai, Z., Eichhardt, I., & Hajder, L. (2018). Accurate calibration of multi-lidar-multi-camera systems. *Sensors*, 18(7), 2139.

Ravi, R., Lin, Y. J., Elbahnasawy, M., Shamseldin, T., & Habib, A. (2018a). Simultaneous system calibration of a multi-lidar multicamera mobile mapping platform. *IEEE Journal of selected topics in applied earth observations and remote sensing*, 11(5), 1694-1714.

Ravi, R., Lin, Y. J., Elbahnasawy, M., Shamseldin, T., & Habib, A. (2018b). Bias impact analysis and calibration of terrestrial mobile lidar system with several spinning multibeam laser scanners. *IEEE Transactions on Geoscience and Remote Sensing*, 56(9), 5261-5275.

Revelles, J., Urena, C., & Lastra, M. (2000). An efficient parametric algorithm for octree traversal.

Riegl. Riegl miniVUX-1DL. Available Online: http://www.riegl.com/products/ unmanned-scanning/riegl-minivux-1dl (accessed on 4 May 2021).

Schonberger, J. L., & Frahm, J. M. (2016). Structure-from-motion revisited. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4104-4113).

Sony. Sony ILCE7RM3/B. Available Online: https://electronics.sony.com/imaging/interchangeable-lens-cameras/fullframe/p/ilce7rm3-b (accessed on 4 May 2021a).

Sony. Sony RX1RM2. Available Online: https://www.sony.com/electronics/cyber-shot-compact-cameras/dsc-rx1rm2/specifications (accessed on 4 May 2021b).

Trimble. Trimble R10 Model 2 GNSS System. Available online: https://geospatial.trimble.com/sites/geospatial.trimble.com/files/2019-04/022516-332A_TrimbleR10-2_DS_USL_0419_LR.pdf (accessed on 4 May 2021).

Velodyne. Velodyne VLP-32C User Manual. Available Online: https://icave2.cse.buffalo.edu/resources/sensor-modeling/VLP32CManual.pdf (accessed on 4 May 2021).

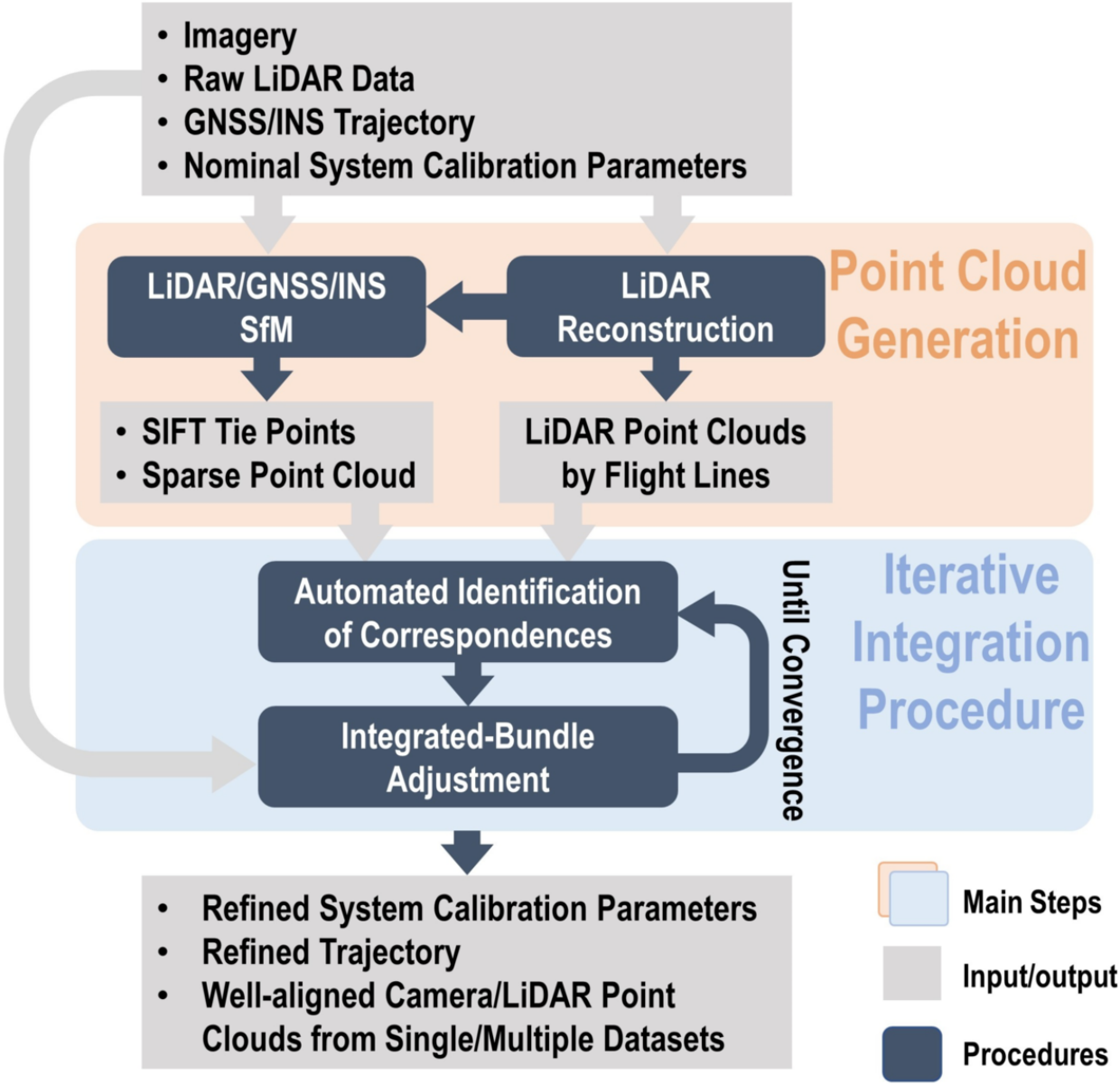
Westoby, M. J., Brasington, J., Glasser, N. F., Hambrey, M. J., & Reynolds, J. M. (2012). 'Structure-from-Motion'photogrammetry: A low-cost, effective tool for geoscience applications. *Geomorphology, 179*, 300-314.

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. Chemometrics and intelligent laboratory systems, 2(1-3), 37-52.

Xie, Y., Shao, R., Guli, P., Li, B., & Wang, L. (2018, June). Infrastructure based calibration of a multi-camera and multi-lidar system using apriltags. *In 2018 IEEE Intelligent Vehicles Symposium (IV)* (pp. 605-610). IEEE.

Zhou, L., Li, Z., & Kaess, M. (2018, October). Automatic extrinsic calibration of a camera and a 3d lidar using line and plane correspondences. *In 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 5562-5569). IEEE.

Zhou, T., Hasheminasab, S. M., Ravi, R., & Habib, A. (2020). LiDAR-aided interior orientation parameters refinement strategy for consumer-grade cameras onboard UAV remote sensing systems. *Remote Sensing*, *12*(14), 2268.


• Imagery
• Raw LiDAR Data
• GNSS/INS Trajectory
• Nominal System Calibration Parameters
LiDAR/GNSS/INS SfM
LiDAR Reconstruction
Point Cloud Generation
• SIFT Tie Points
• Sparse Point Cloud
LiDAR Point Clouds by Flight Lines
Automated Identification of Correspondences
Integrated-Bundle Adjustment
Until Convergence
Iterative Integration Procedure
• Refined System Calibration Parameters
• Refined Trajectory
• Well-aligned Camera/LiDAR Point Clouds from Single/Multiple Datasets
Main Steps
Input/output
Procedures