

Article

GNSS/INS-Assisted Structure from Motion Strategies for UAV-Based Imagery over Mechanized Agricultural Fields

Seyyed Meghdad Hasheminasab ¹, Tian Zhou ¹ and Ayman Habib ^{1,2,*}

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

² Civil Engineering Center for Applications of UAS for a Sustainable Environment (CE-CAUSE), Lyles School of Civil Engineering, Purdue University, West Lafayette, IN 47907, USA

* Correspondence: ahabib@purdue.edu; Tel.: +1-765-496-0173

Received: 20 December 2019; Accepted: 19 January 2020; Published: 21 January 2020



Abstract: Acquired imagery by unmanned aerial vehicles (UAVs) has been widely used for three-dimensional (3D) reconstruction/modeling in various digital agriculture applications, such as phenotyping, crop monitoring, and yield prediction. 3D reconstruction from well-textured UAV-based images has matured and the user community has access to several commercial and opensource tools that provide accurate products at a high level of automation. However, in some applications, such as digital agriculture, due to repetitive image patterns, these approaches are not always able to produce reliable/complete products. The main limitation of these techniques is their inability to establish a sufficient number of correctly matched features among overlapping images, causing incomplete and/or inaccurate 3D reconstruction. This paper provides two structure from motion (SfM) strategies, which use trajectory information provided by an onboard survey-grade global navigation satellite system/inertial navigation system (GNSS/INS) and system calibration parameters. The main difference between the proposed strategies is that the first one—denoted as partially GNSS/INS-assisted SfM—implements the four stages of an automated triangulation procedure, namely, imaging matching, relative orientation parameters (ROPs) estimation, exterior orientation parameters (EOPs) recovery, and bundle adjustment (BA). The second strategy—denoted as fully GNSS/INS-assisted SfM—removes the EOPs estimation step while introducing a random sample consensus (RANSAC)-based strategy for removing matching outliers before the BA stage. Both strategies modify the image matching by restricting the search space for conjugate points. They also implement a linear procedure for ROPs' refinement. Finally, they use the GNSS/INS information in modified collinearity equations for a simpler BA procedure that could be used for refining system calibration parameters. Eight datasets over six agricultural fields are used to evaluate the performance of the developed strategies. In comparison with a traditional SfM framework and Pix4D Mapper Pro, the proposed strategies are able to generate denser and more accurate 3D point clouds as well as orthophotos without any gaps.

Keywords: structure from motion; automatic aerial triangulation; unmanned aerial vehicles; GNSS/INS-assisted mapping; image matching; epipolar geometry

1. Introduction

Feeding the growing population will be one of the most challenging tasks for agriculture in the near future [1]. Moreover, food security is increasingly threatened by several factors, such as food waste, climate change, and plant disease [2]. In fact, there is a need to increase the food production

by more than 50% before 2050 in order to feed roughly 10 billion people, almost 3 billion more than the current population [3]. Digital agriculture, also known as smart farming, aims at increasing food production by enhancing decision making through the analysis of digitally collected data and/or information over agricultural fields [4]. In this regard, remotely sensed data acquired by satellite [5], manned aircraft [6], and, most recently, unmanned aerial vehicle (UAV) [7] platforms, are among the most popular digital data sources that are starting to replace many of the traditional in-field manual trait measurements [8]. The easier deployment and possibility of equipping UAVs with a variety of advanced imaging sensors, along with their capability to collect high temporal and spatial resolution data have increased the use of UAVs for digital agriculture, in applications such as phenotyping [9–13], crop monitoring [14–17], and yield estimation/prediction [18–20]. For most of these applications, accurate georeferenced three-dimensional (3D) point clouds and orthophotos are the main products required for various plant trait measurements, such as canopy cover [21], plant height [22], and plant count [23,24].

Structure from motion (SfM), a photogrammetric and computer vision framework, has been widely employed for 3D reconstruction from UAV-based imagery. In most state-of-the-art SfM-based mapping strategies, the final 3D model is reconstructed through four main steps, namely, image matching, relative orientation parameters (ROPs) estimation, exterior orientation parameters (EOPs) recovery, and bundle adjustment (BA). First, using feature extraction algorithms such as scale invariant feature transform (SIFT) [25] or speeded-up robust features (SURF) [26], local features are detected, characterized through descriptors, and matched among overlapping images. In the next step, using the essential matrix and epipolar geometry, ROPs between stereo images are estimated using identified conjugate points. It should be noted that, due to illumination differences, view-point change, and/or repetitive patterns in the images, there are always outliers among the initial corresponding features. As the quality of derived ROPs depends heavily on the correctness of the matched points, the ROP estimation step is usually augmented with an outlier removal strategy, such as the random sample consensus (RANSAC) procedure [27]. Once the ROPs are recovered, in the third step, a local coordinate system is defined and EOPs of the involved imagery as well as 3D coordinates of the matched features are estimated. Finally, a bundle adjustment procedure is conducted to refine the derived EOPs and coordinates of object points.

The reconstructed 3D model can be georeferenced using either ground control points (GCPs) or trajectory information provided by a survey-grade global navigation satellite system/inertial navigation system (GNSS/INS) unit onboard the UAV. The former, known as indirect georeferencing, is the traditional way of georeferencing and is used in most state-of-the-art SfM frameworks, while the latter, known as direct georeferencing, or more specifically, integrated sensor orientation (ISO), is still an active area of research. System calibration of UAV-based, GNSS/INS-assisted imaging systems is a vital step for delivering accurately georeferenced products through direct georeferencing. Such system calibration considers both spatial and temporal aspects. Spatial system calibration parameters include internal characteristics of the onboard camera(s), known as interior orientation parameters (IOPs), as well as mounting parameters which describe the differences in the position and orientation between the GNSS/INS body frame and camera(s) frame [28–30]. On the other hand, temporal system calibration aims at solving and correcting for any possible time delay in the synchronization between the GNSS/INS unit and the camera(s) onboard the UAV system [31,32].

Currently, a wide range of commercial—e.g., Pix4D Mapper Pro, PhotoScan, and DroneDeploy—and opensource tools—e.g., OpenDroneMap, Regard3D, Meshroom, and COLMAP—have automated the SfM process. Furthermore, several studies proposed transparent techniques for image-based 3D reconstruction [33–35]. However, the majority of these approaches exhibit poor performance in some applications, such as digital agriculture, coastal monitoring, and urban area mapping due to poor and/or repetitive patterns in the acquired imagery. These repetitive patterns are mainly caused by crop nature and mechanized planting in agricultural fields (see Figure 1). This fact highlights the need for developing a robust and transparent automated aerial triangulation framework for UAV-based

mapping, which is capable of handling images with challenging texture conditions, especially those over mechanized agricultural fields.

This paper presents new SfM strategies for aerial triangulation of UAV-based imagery to overcome the above-mentioned limitations of existing approaches by fully exploiting the information provided by the onboard GNSS/INS unit. More specifically, the presented strategies aim at mitigating the impact of repetitive texture pattern on image-based 3D reconstruction using the GNSS/INS trajectory. In the proposed approaches, using system calibration parameters and GNSS/INS information, the four steps of the traditional SfM framework are modified as follows:

- **Image matching:** Rather than conducting a traditional exhaustive search among the feature descriptors within the images of a stereo-pair, GNSS/INS information is used to reduce the search space. This can mitigate some of the matching ambiguity problems caused by repetitive patterns.
- **Relative orientation parameter estimation:** GNSS/INS-based ROPs are used as initial values in an iterative ROP estimation and outlier removal procedure.
- **Exterior orientation parameter recovery:** For this step, two strategies, denoted as partially GNSS/INS-assisted SfM and fully GNSS/INS-assisted SfM, are employed. The first strategy implements a traditional incremental EOP recovery, which derives the image EOPs in a local coordinate system while removing matching outliers. In the second strategy, however, as the georeferencing parameters can be directly derived from the GNSS/INS information, the EOP recovery step is removed from the SfM framework.
- **Bundle adjustment:** A GNSS/INS-assisted bundle adjustment is conducted to refine the derived object points, camera position and orientation parameters, and/or system calibration parameters. Also, in the fully GNSS/INS-assisted SfM framework, the bundle adjustment step is augmented with a preceding RANSAC strategy for removing matching outliers that were not detected through prior steps.



Figure 1. A sample image from a mechanized agricultural field. The similarity between plants along/across the rows causes repetitive patterns in the captured images.

The rest of the paper is organized as follows: Section 2 focuses on related works, Section 3 describes the UAV-based imaging system and data used in this study, Section 4 introduces the proposed SfM frameworks, Section 5 presents the experimental results, and finally, Section 6 provides conclusions and recommendations for future work.

2. Related Works

In this section, a literature review of existing research efforts towards automated aerial triangulation is given. More specifically, the image matching strategies along with feature detection/characterization algorithms are first introduced, then, various approaches for ROP estimation are reviewed, and lastly, two strategies for exterior orientation recovery—incremental and global—are discussed. As the various bundle adjustment techniques have been sufficiently reviewed in previous studies [33,36], this step is not discussed further.

2.1. Image Matching

The first and foremost processing step in SfM is the identification of features in individual images. Extracted features are then matched in overlapping images. The most popular strategy follows a detect-and-describe framework that identifies a set of interest points, and generates descriptors based on local regions around the extracted interest points. Traditional feature detector and descriptor approaches—such as features from accelerated segment test (FAST) [37], SIFT [25], and SURF [26]—have been thoroughly evaluated [38–40]. Recently, with the rapid development of deep learning strategies, various convolutional neural network-based approaches are proposed to conduct interest point detection and descriptor learning [41–43]. Among the above-mentioned approaches, SIFT is still the most widely-used feature detection and characterization algorithm since it is invariant against several factors, such as scaling, rotation, and illumination changes. Recently, graphics processing units (GPUs) have been adopted in the implementation of SIFT-based algorithms [44,45] to improve the computational efficiency. The original implementation of the SIFT algorithm extracts a large collection of features and generates 128-dimensional feature descriptors. Then, for a given feature, p_1 , extracted in the left image, its conjugate feature in the right image is identified in three steps: (1) the L2 distances between descriptors of p_1 and all candidate features in the right image are computed, (2) if the ratio between the closest and second closest distances is smaller than a given threshold (e.g., 0.8), the feature with closest distance, p_2 , is accepted as a conjugate feature, and (3) steps 1 and 2 are used to identify the conjugate feature for p_2 from potential matching features in the left image, wherein the pair (p_1, p_2) is considered to be valid if left-to-right and right-to-left matchings are consistent.

2.2. Estimation of Relative Orientation Parameters

Once initial matches among overlapping images are derived, ROPs that describe the relative rotational and translational relationship (R, \vec{r}) between the images of a stereo-pair are estimated. The involved parameters in ROP recovery consist of three rotation angles and two translation parameters [46], since the translational relationship between two images can only be estimated up to an arbitrary scale due to the lack of knowledge about dimensions in the actual scene. The fundamental mathematical model for ROP recovery is the co-planarity constraint [47], which enforces the fact that two light rays connecting the perspective centers of the imaging sensor, object point, and the respective image points should lie on the same plane. The coplanarity constraint is mathematically introduced in Equation (1), where two conjugate image points corresponding to a given object point are denoted as p_1 and p_2 and the cross-product operation is denoted as “ \times ”. Since one is dealing with a non-linear model in the involved unknowns, an iterative least-squares adjustment procedure should be used starting from approximate values of the unknowns. Under certain circumstances, coming up with meaningful approximate values could be quite challenging. To relax this requirement, several closed-form solutions, which are linear, have been developed for ROP recovery. These approaches are based on the essential matrix introduced by the computer vision community [46], while assuming that the cameras are calibrated. The essential matrix, E , is a 3×3 matrix, that relates corresponding points in a stereo-pair, as expressed by Equation (2). The essential matrix is represented by the rotation matrix, R , and translation vector, \vec{r} , as in Equation (2), where \hat{r} is a skew-symmetric matrix corresponding to \vec{r} , which is used to replace the cross product by a matrix-vector product. The nine elements of the essential matrix must satisfy four constraints, which are defined as follows:

1. The elements of the essential matrix can be only determined up to a scale,
2. The essential matrix should have a rank of two, thus the determinant of the matrix should be zero, and
3. Two trace constraints [46], as presented in Equation (3), should be satisfied.

$$p_1 \cdot (\vec{r} \times Rp_2) = 0; \quad (1)$$

$$p_1^T E p_2 = 0, \text{ where } E = \hat{r}R; \quad (2)$$

$$EE^T E - \frac{1}{2} \text{trace}(EE^T) E = 0. \quad (3)$$

Motivated by the essential matrix, Longuet-Higgins [47] proposed a simple eight-point algorithm for relative orientation estimation by solving a set of linear equations without the need for prior information. This approach will fail under certain ‘degenerate’ eight-point configurations, such as more than four points lying on a straight line, or more than seven points lying on a plane. Besides the ‘degenerate’ point configuration issue, the eight-point algorithm is extremely susceptible to noise. Hartley [48] examined the eight-point algorithm and confirmed that the bad performance can be traced to implementations that do not take numerical considerations into account when solving the linear equations. Hartley proved that the performance can be significantly improved by adopting a simple normalization (translation and scaling) of the coordinates of the matched points. Hartley also validated, through thousands of experiments, that the modified eight-point algorithm can achieve similar results when compared to complicated iterative algorithms [49,50]. However, these eight-point algorithms do not consider the inherent constraints when solving for the elements of the essential matrix. Thus, the eight-point algorithms are mainly suited for non-calibrated cameras.

Given that the essential matrix has five degrees of freedom, five-point algorithms have been proposed. Among various five-point algorithms, the one developed by Nistér [51] has the best efficiency. Such an algorithm is based on computing the coefficients of a tenth-degree polynomial and sequentially finding its roots with the help of Gauss–Jordan elimination. Nistér also stated that this five-point algorithm should be used in conjunction with a RANSAC strategy to remove matching outliers. Based on Nistér’s five-point algorithm, Li and Hartley [52] proposed a simpler version for ROP estimation. By adopting a hidden variable resultant technique [53], their new five-point algorithm eliminated many unknowns at once instead of the traditional sequential elimination.

In some specific cases, prior information related to the trajectory position and/or orientation of the system is available. Several approaches take this information into consideration for specific applications. For example, He and Habib [54] developed a three-point closed-form solution for automated motion parameter estimation of a multi-camera indoor mapping system while considering a planimetric translation and single rotation around the vertical axis, thus leading to three independent parameters being estimated (i.e., one rotation angle and two translation parameters). In their work, He and Habib used the known spatial distance between the elements of the multi-camera system to estimate the absolute planimetric translation. Ortin and Montiel [55] proposed a two-point algorithm for indoor robot motion under the same assumption as the three-point approach proposed by He and Habib [54]. However, only two parameters were estimated since the translation relation can only be determined up to a scale. Scaramuzza et al. [56] proved that by exploiting the nonholonomic constraints of a wheeled vehicle, it is possible to parameterize the motion with only one feature correspondence. Similarly, Hoang et al. [57] proposed a one-point algorithm for planar motion estimation using a monocular omnidirectional camera based on an extended Kalman filter.

With regard to UAV applications, He and Habib [58] presented two strategies for reliable estimation of ROPs in the presence of a high percentage of outliers. Assuming that the UAV platform is moving at a constant flying height while maintaining the camera in a nadir-looking orientation, a two-point closed-form solution is derived as the first strategy. The two-point solution can deal with stereo pairs that have any heading angle and planimetric translation. However, it cannot tolerate significant variations in the tilt angles as well as flying height differences between the stereo images. The proposed two-point solution was integrated within a RANSAC framework to remove matching outliers. The second strategy, which is denoted as the iterative five-point algorithm, starts from prior information regarding the flight trajectory to define a linearized model to estimate a refined set of ROPs. More specifically, the co-planarity model is simplified to a linear equation involving the corrections to the approximate values for the ROPs as well as the image coordinates of conjugate point pairs. The corrections are then used to refine the approximate ROPs. This process is iteratively executed while removing potential outliers until a convergence criterion is achieved. Based on these

two strategies, He et al. [33] proposed a framework for automated aerial triangulation using UAV images. In the ROP recovery step of their proposed work, a hybrid strategy, which integrates both the two-point and iterative five-point algorithms, was used. In such a strategy, the two-point approach is first conducted to provide initial ROP estimates, then, the derived parameters are further refined through the implementation of the iterative five-point approach.

2.3. Estimation of Exterior Orientation Parameters

EOP recovery deals with the estimation of the camera position/orientation for all the images in the block relative to a single, arbitrary reference frame. The initial recovery of EOPs for the involved images within the established stereo-pairs is achieved through either incremental or global strategy. The incremental strategy normally starts from a stereo-pair to define the datum, and then estimate the EOPs for the remaining images relative to that datum through an augmentation process. For example, Snavely et al. [59] reconstructed 3D models from internet image datasets using an incremental strategy. A single pair of images with a large number of matches as well as large baseline was first chosen to define the datum. Then, other images were added to the model one by one where their EOPs were estimated using 3D points that have been reconstructed earlier. The reconstruction quality of this approach heavily depends on the selection of the first image pair and the order of adding the remaining images. In order to solve this problem, Dunn and Frahm [60] developed a hierarchical uncertainty-driven model to sequentially find the best image to add to the sets of images that have been referenced relative to the datum. This procedure results in maximum reduction of the model's 3D uncertainty. Another strategy to estimate the EOPs for each augmented image is through single rotation averaging followed by translation averaging. A least-squares approach that estimates the nine unknown elements of the rotation matrix can be derived through single rotation averaging [61]. An alternative linear solution for the single rotation averaging could be achieved through quaternions [62]. However, both of these approaches fail to consider the inherent constraints within a rotation (i.e., the orthogonality constraints for a rotation matrix and the unit length constraint for a quaternion rotation). He et al. [33] proposed a new quaternion-based approach while enforcing the unit length constraint for the quaternion rotation. In terms of estimating positional components, He et al. [33] proposed two different closed-form solutions to handle images captured either in a block or a linear trajectory configuration. As an alternative to sequentially adding one image at a time to the referenced images, there are some incremental approaches [63,64] that hierarchically reconstruct different sub-models and then merge them into the final model. In general, the main drawback of incremental algorithms is significant drifting errors when dealing with large image blocks [65]. Adopting a frequent, intermediate bundle adjustment routine can help reduce drifting errors but will create a computational bottleneck.

Unlike the incremental strategy, global approaches estimate EOPs of all images simultaneously and have better performance in terms of efficiency and accuracy. Global approaches are usually conducted in two steps to sequentially solve for the camera rotations and positions. First, a multiple rotation averaging algorithm is used to estimate the rotations of all images. Then, with the help of estimated rotations, global translation averaging is used to estimate positional components of the EOPs. More specifically, given a set of relative rotations, R_j^i , between the coordinate frames of images i and j , multiple rotation averaging aims at finding the optimum global rotation estimates for all involved images (e.g., R_i^{global} , R_j^{global}) while minimizing the difference between R_j^i and $(R_i^{global})^T R_j^{global}$ according to a prespecified metric. The difference metrics include geodesic, quaternion, and chordal [61]. Govindu [66] suggested a quaternion averaging strategy for multiple rotations while assuming a uniform, Gaussian distribution for the rotation error. Thus, the rotation averaging problem can be solved using a maximum likelihood estimation strategy. More specifically, a closed-form linear least-squares solution was derived using singular value decomposition (SVD). Martinec and Pajdla [62] introduced a rotation averaging strategy using the chordal metric and compared their approach with the linear quaternion averaging, the obtained rotation result

from Martinec and Pajdla [62] does not satisfy the orthogonality constraints of a rotation matrix. Instead of using Euler angles or quaternions, recent research efforts, which are based on Lie-algebra representations and robust L1 optimization, have demonstrated better performance when solving the multiple rotation averaging problem [67]. Considering that the aforementioned approaches are complex to implement, He et al. [33] proposed a simple global rotation averaging algorithm by converting the mathematical relationship among R_j^{global} , R_i^{global} , and R_j^i (i.e., $R_j^{global} = R_i^{global} R_j^i$) into a linear model while ignoring the inherent orthogonality constraints. Assuming that there are m available stereo-pairs within a set of n overlapping images, a system of $9m$ equations in $9n$ unknown parameters (i.e., nine elements within each unknown rotation matrix) can be established. A closed-form solution can be derived through SVD.

Translation averaging evaluates camera positions, typically with their rotation parameters estimated beforehand. It is impossible to determine the absolute camera positions from ROPs or essential matrix as the translation vectors among different image-pairs are only determined up to an arbitrary scale. Existing research efforts for global translation estimation include linear approaches which are mainly based on consistency constraints among different translation vectors relating stereo images within a block [66]. However, such approaches cannot deal with images captured along a linear-trajectory configuration. Different from Reference [66], Sinha et al. [68] determined the global translations through stereo-based registration with the help of 3D points reconstructed in all possible stereo-pairs. Arie-Nachimson et al. [69] introduced another solution for global translation estimation while using a novel decomposition of the essential matrix. However, this solution still suffers from the degeneracy caused by linear trajectory configuration. In order to resolve such degeneracy, Cui et al. [70] utilized corresponding image points, which are derived through a feature tracking process, to establish a common scale for all the translation parameters. Similarly, He et al. [33] proposed a linear model which could be solved by SVD while considering both ROPs and involved matched tracked points. Despite its advantages in terms of efficiency and accuracy, global SfM approaches are fragile and sensitive to the noise level caused by poor ROP estimation arising from matching outliers. Thus, incremental SfM is still the most commonly used strategy in existing commercial software packages.

3. Data Acquisition System Specifications and Configurations of Case Studies

In this section, we introduce the platform and imaging system used in this research, including sensor specifications and system calibration strategy. Then, the acquired datasets are described in terms of flight configuration and the corresponding availability of ground control information for accuracy evaluation of derived products.

3.1. Data Acquisition System

In order to validate the feasibility of the proposed research, a custom-built, UAV-based mobile mapping system is employed for data acquisition over agricultural fields. This system consists of a Dà-Jiāng Innovations (DJI) Matrice 600 Pro (M600P) carrying a Sony α 7R III (ILCE-7RM3) Red-Green-Blue (RGB) camera, a Velodyne Puck Lite LiDAR sensor, and Trimble APX-15 UAV v3 GNSS/INS unit (shown in Figure 2a). LiDAR data is only utilized for the evaluation of image-based reconstruction quality in the absence of GCPs. The APX-15 unit provides direct georeferencing information—i.e., the position and orientation of the inertial measurement unit (IMU) body frame at 200 HZ. After post-processing of the GNSS/INS data, the expected positional accuracy is around 2 to 5 cm, and the accuracy for pitch/roll and heading is 0.025° and 0.08° , respectively. The Sony α 7R III camera is a 42 megapixel (MP) camera with a 7952×5304 complementary metal oxide semiconductor (CMOS) array, $4.5 \mu\text{m}$ pixel size, and a lens with 35 mm nominal focal length. The internal characteristics of the camera are estimated using the United States Geological Survey (USGS) simultaneous multi-frame analytical calibration (SMAC) distortion model, in a calibration procedure similar to the one proposed in Reference [71]. In this study, estimated IOPs include the principal distance c , principal point coordinates (x_p, y_p) , and four radial and de-centering lens distortion parameters (K_1, K_2, P_1, P_2) . The camera is

triggered at a frame interval of 1.5 s, using an Arduino Micro development board. Each image is also time-tagged using a direct feedback synchronization approach, where the camera flash hot-shoe is utilized to generate a signal at the time of camera exposure. This camera feedback signal is sent to the GNSS/INS unit and a corresponding event marker is recorded in the GNSS/INS trajectory.

As mentioned earlier, system calibration is a key step towards achieving accurately georeferenced products from GNSS/INS-assisted mobile mapping systems. In this research, the mounting parameters, defined by the boresight angles and lever arm components, between the GNSS/INS unit and other onboard sensors (i.e., LiDAR and RGB/hyperspectral cameras) are rigorously determined using the calibration strategy introduced in References [28,29]. In this strategy, the mounting parameters are simultaneously estimated through minimizing the discrepancies among linear/planar features and conjugate points extracted from LiDAR point clouds and images from different flight lines. Also, similar to the approach proposed in Reference [31], a time offset calibration is done to solve and correct for any possible time delay between the actual camera exposure and recorded event marker by the GNSS/INS unit. This approach modifies the collinearity equations so that the time delay can be directly estimated in the bundle adjustment process. The positional accuracy of the image/LiDAR-based products using this as well as similar systems have been extensively investigated through several studies [29,33,72,73].



Figure 2. The unmanned aerial vehicle (UAV)-based mobile mapping system used in this study: (a) the sensors mounted on the UAV and (b) the orientation of the global navigation satellite system/inertial navigation system (GNSS/INS) body frame (yellow), laser unit frame (green), and RGB camera frame (red) coordinate systems.

3.2. Dataset Description

In this study, eight datasets collected over six agricultural fields are used to evaluate the performance of the proposed approaches. These agricultural fields are used for seed breeding trials and involve different crops and plots with different sizes, planting densities, and genotypes. The datasets were acquired during the summer of 2019 as a part of the Purdue's Transportation Energy Resources from Renewable Agriculture (TERRA) project. All the data collection sites are located in Indiana, USA. As shown in Figure 3, the six sites are as follows: two agricultural fields at Purdue's Agronomy Center for Research and Education (ACRE), denoted as ACRE-42 and ACRE-21C, one at Romney, one at Windfall, and two fields at Atlanta, denoted as Atlanta-1 and Atlanta-2. For all the conducted flight missions, the DJI GS Pro mission planning software was used for autonomous flight path programming. Flight configurations and crop types for the different datasets are summarized in Table 1.



Figure 3. Location of the surveyed agricultural fields: Purdue’s Agronomy Center for Research and Education (ACRE) (including ACRE-42 and ACRE-21C), Romney, Windfall, and Atlanta, including two fields, Atlanta-1 and Atlanta-2.

Table 1. Flight configurations and crop designation for the different datasets.

Acquisition Date	Field	Crop	Flying Height (m)	Ground Speed (m/s)	Lateral Distance ¹ (m)	GSD ² (cm)	Overlap/Side-Lap (%)	# ³ of Images
20190810	ACRE-42	Sorghum	47	5.0	13.0	0.76	80/78	562
20190905	ACRE-42	Sorghum	47	5.0	13.0	0.76	80/78	555
20190823	ACRE-21C	Popcorn	47	5.0	8.5	0.76	80-85	215
20190904	ACRE-21C	Popcorn	44	4.1	9.0	0.71	84/84	217
20190724	Romney	Popcorn	45	6.0	14.0	0.72	76/76	846
20190718	Windfall	Maize	45	6.5	14.0	0.72	75/76	414
20190802	Atlanta 1	Maize	45	5.5	13.0	0.72	78/77	254
20190802	Atlanta 2	Maize Soybeans	45	5.5	13.0	0.72	78/77	355

¹ Lateral distance between neighboring flight lines. ² Ground sampling distance. ³ Number.

A total of eight and sixteen checkerboard targets, used as check points, were deployed at the perimeter of ACRE-21C and ACRE-42 fields, respectively. Figure 4 shows aerial views of these two fields with enhanced representation of the checkerboard targets. The ground coordinates of these checkerboard targets were surveyed through the real-time kinematic (RTK) technique using a Trimble R10 GNSS receiver, with an advertised horizontal accuracy of 8 mm + 1 ppm, and vertical accuracy of 15 mm + 1 ppm, where the term “ppm” stands for parts per million of the distance between the rover receiver and the GNSS base-station. In the employed survey with a 6 km baseline, the expected horizontal and vertical accuracy from the R10 is in the range of 2 to 3 cm and 3 to 4 cm, respectively. The checkerboard targets were then identified in the images to assess the 3D reconstruction accuracy. There were no checkerboard targets available for the remaining four study sites. It should be noted that in comparison with ACRE-42 and ACRE-21C test fields, the Romney, Windfall, Atlanta-1, and Atlanta-2 experimental fields include plots with larger size, higher planting density, and more similar genotypes. This fact induces more challenging texture conditions for images captured over those fields. Figure 5 shows some sample images captured by the Sony α 7R III camera over the six agricultural test fields.

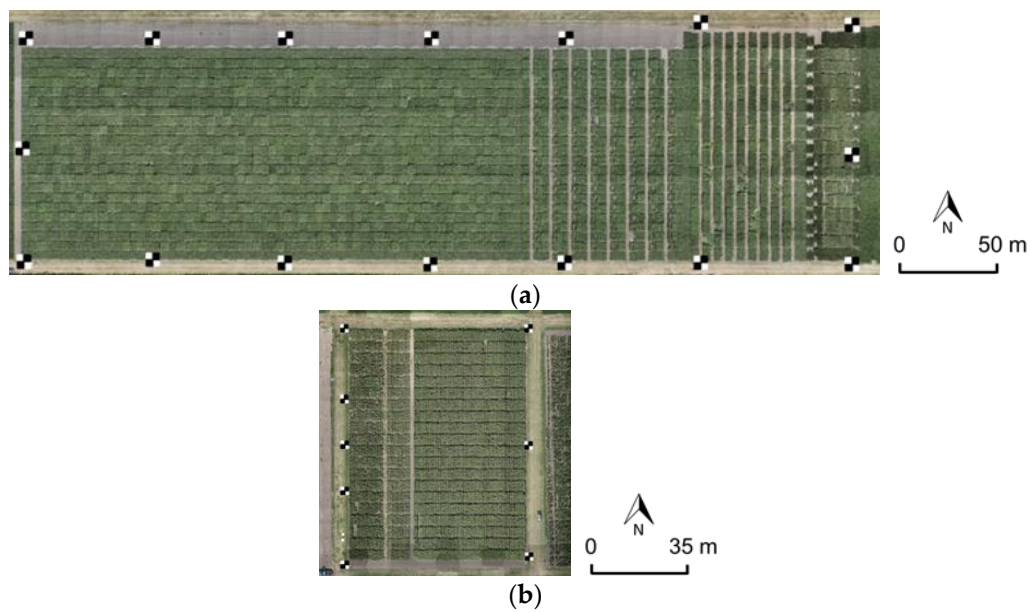


Figure 4. Test fields with enhanced representation of checkerboard targets: (a) ACRE-42 and (b) ACRE-21C.

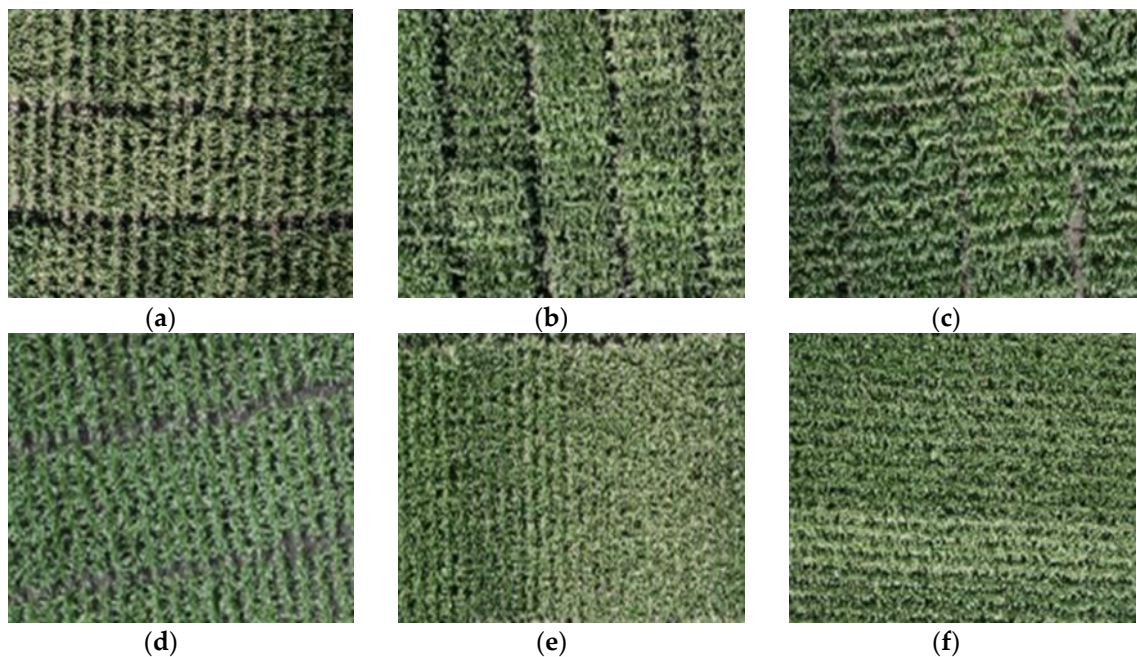


Figure 5. Sample images from: (a) ACRE-21C, (b) ACRE-42, (c) Romney, (d) Windfall, (e) Atlanta-1, and (f) Atlanta-2 (repetitive patterns are more pronounced in c, d, e, and f).

4. Methodology

In this section, the proposed framework, which takes advantage of the available GNSS/INS trajectory to facilitate the 3D reconstruction process, is introduced. In order to improve the image-based 3D reconstruction, the proposed framework uses the GNSS/INS trajectory in the image matching, ROP estimation, and bundle adjustment steps of the SfM framework. Figure 6 illustrates the workflow of the proposed approaches. The first block in Figure 6 corresponds to stereo image matching. Here, the SIFT algorithm is first applied on all images to detect and extract local features along with their descriptors. Then, in order to mitigate the matching ambiguity caused by similar feature descriptors, trajectory information is used to reduce the search space for the identification of conjugate features.

The second block deals with ROP estimation wherein the relative orientation parameters for the different stereo-pairs are initially derived using the GNSS/INS trajectory and system calibration parameters. These ROPs are then refined through the adoption of the iterative five-point approach, where matching outliers are detected and removed. Once the ROPs of the stereo-pairs are derived and some of the matching outliers are removed, in the next step, two strategies are introduced, denoted as partially GNSS/INS-assisted SfM and fully GNSS/INS-assisted SfM. In the partially GNSS/INS-assisted SfM, a traditional incremental approach is implemented to derive the image EOPs and remove more matching outliers, followed by a GNSS/INS-assisted bundle adjustment. In the fully GNSS/INS-assisted SfM, inspired by the fact that EOPs are already available from the GNSS/INS trajectory, first, a RANSAC strategy for further removal of matching outliers is applied, and then, a GNSS/INS-assisted bundle adjustment is conducted to simultaneously refine the GNSS/INS trajectory, 3D coordinates of matched points, and/or system calibration parameters. As described above, the partially GNSS/INS-assisted SfM has a four-step framework, while the fully GNSS/INS-assisted SfM is conducted in three main steps. In the remainder of this section, the proposed strategies for image matching, ROP estimation, and bundle adjustment steps are discussed in detail. Details of the implemented EOP recovery step can be found in Reference [33].

4.1. Stereo Image Matching

In the first step of the proposed SfM framework, the SIFT detector and descriptor algorithm is applied on the entire image network. The traditional matching strategy finds conjugate points between two overlapping images, hereafter denoted as left and right images, by comparing each feature descriptor in the left image with all feature descriptors in the right images. This process is depicted in Figure 7. More specifically, given a feature descriptor in the left image, its Euclidian distances to descriptors for all the features in the right image are computed, and if the nearest distance is significantly less than the second nearest distance, a matching hypothesis is established. Finally, as described in Section 2.1, a left-to-right and right-to-left consistency check is applied to remove potential outliers from the identified matches. As mentioned earlier, in some applications, such as digital agriculture, repetitive patterns in the imagery cause very high similarity between the feature descriptors, thus resulting in fewer matches and/or matches with a high percentage of outliers. To overcome this deficiency, the proposed matching strategy exploits the GNSS/INS trajectory and ground elevation information, with the latter derived from the mission planning parameters, to reduce the search space for potential matches.

In this study, the positional component of the GNSS/INS trajectory is used for defining the candidate stereo-pairs for image matching and ROP estimation. More specifically, given an image, its K nearest images are selected as its candidate pairs, where the value of K is selected according to the a-priori-known percentage of overlap and side-lap between the images. For each candidate stereo-pair, feature matching is conducted in a forward-backward projection strategy. In this regard, based on the well-known collinearity equations while using the available camera IOPs, trajectory-derived EOPs, and an approximate ground elevation, each feature in the left image is first projected to the object space, and then back-projected onto the right image. Therefore, given a feature in the left image, an approximate location of its conjugate point in the right image is estimated, as illustrated in Figure 8a. The predicted point in the right image is then used to define a search window with a user-defined size, as shown in red on the right image in Figure 8b. The search window size can be determined according to the accuracy of the trajectory information and approximate ground elevation as well as GSD of the imagery. In the next step, the search space for potential matches (and consequently matching ambiguity) is further reduced through deriving an epipolar line, shown in green in Figure 8b, in the right image for each feature in the left image using the GNSS/INS trajectory. Then, among all SIFT features in the right image, only those which are located inside the search window as well as inside a buffer (shown in blue in Figure 8b) around the epipolar line are considered as potential conjugate features. It is worth mentioning that although the matching search space is reduced in the proposed

algorithm, conducting the forward-backward projection strategy for each extracted feature could make the proposed algorithm more computationally expensive than the traditional matching approach.

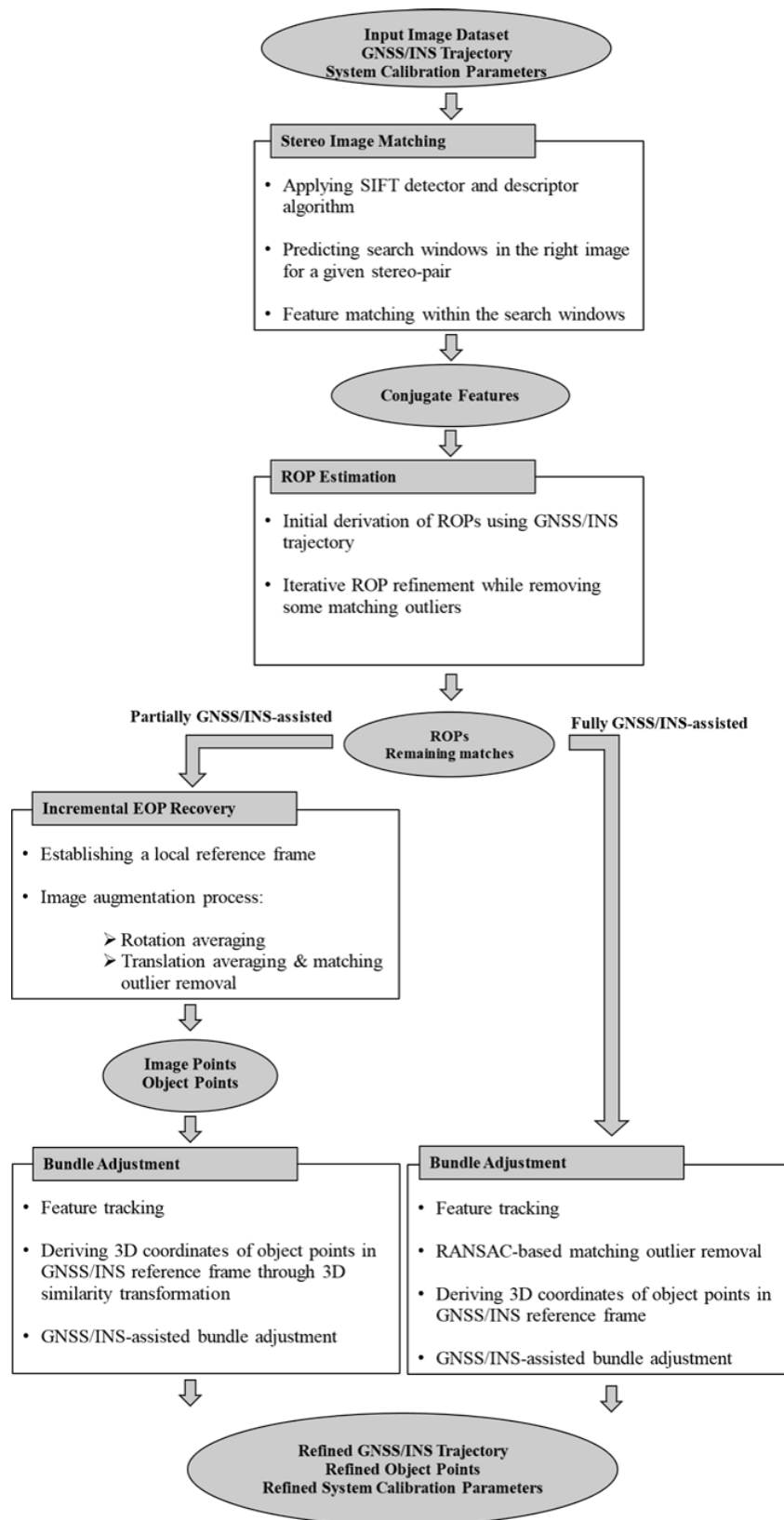


Figure 6. Workflow of the proposed GNSS/INS-assisted structure from motion (SfM) strategies.

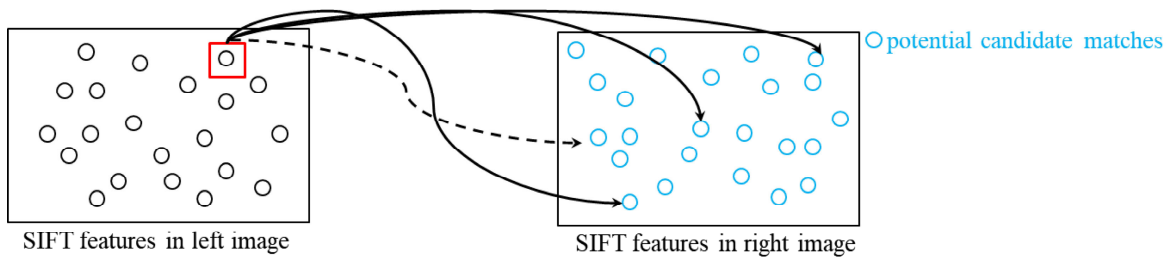


Figure 7. Traditional matching strategy: descriptor similarity evaluation is conducted between the selected left feature and all right features.

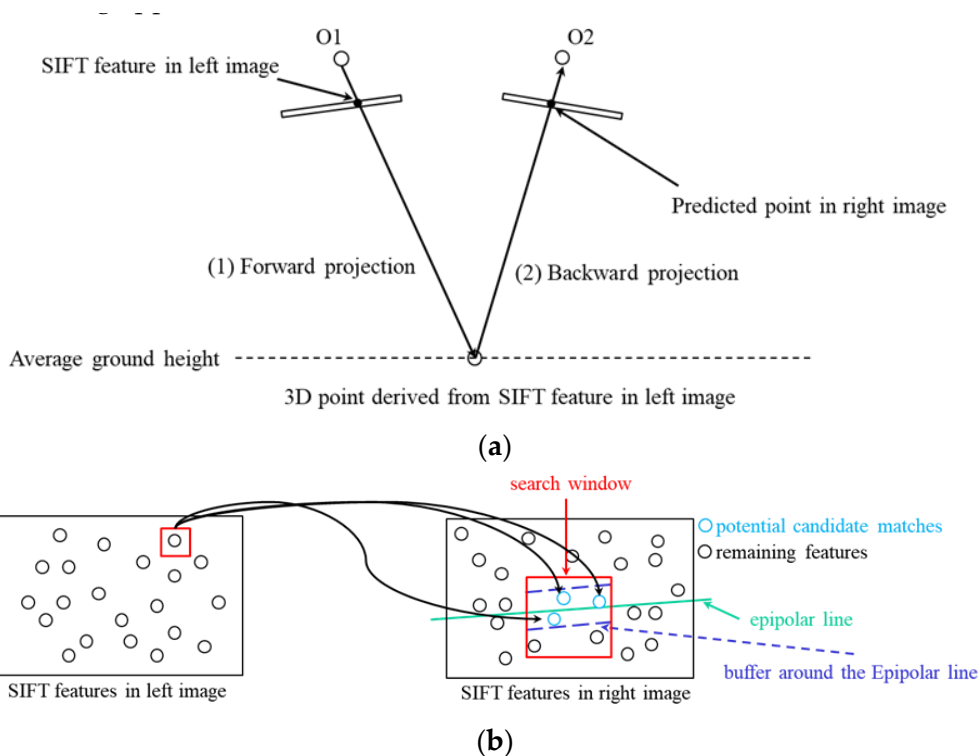


Figure 8. The proposed matching strategy: (a) forward-backward projection and (b) stereo matching.

In the next step, similar to traditional matching strategies, a similarity evaluation based on the Euclidean distances between each left feature descriptor and their potential conjugate features descriptor in the right image is conducted to establish the matching hypothesis. It should be noted that the matching process for a selected feature in the left image is considered successful only when the ratio between the distance to the closest descriptor (d_1) and the distance to the second closest descriptor (d_2) is smaller than a predefined threshold (i.e., $d_1/d_2 < \delta$). Finally, a left-to-right and right-to-left consistency check is conducted for an initial removal of obvious matching outliers. Figure 9 shows sample matching results from the traditional and proposed matching strategies. As expected, the proposed approach leads to more matches when compared to the traditional strategy. Moreover, those matches seem to have less percentage of outliers.

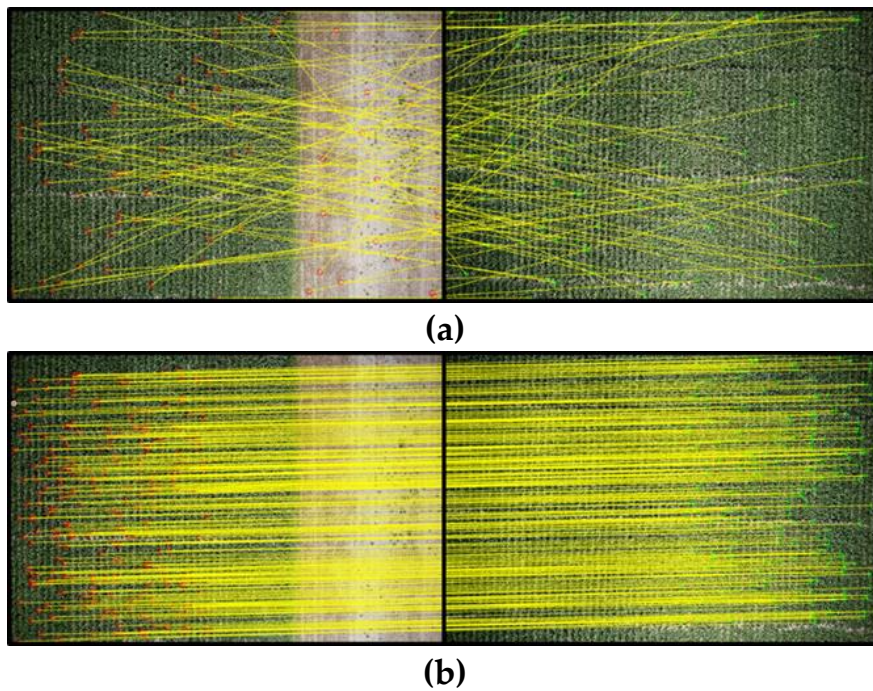


Figure 9. Sample matching results superimposed on a stereo pair from Atlanta-1 dataset: (a) traditional matching with 83 matches and (b) proposed strategy with 268 matches.

4.2. Automated Relative Orientation

Accurate estimation of ROPs is a prerequisite for image-based 3D reconstruction. Considering an arbitrary scale for a stereo-pair, ROPs consist of five parameters, including three rotation angles and two components of the translation vector between two camera stations. As mentioned earlier, the two major strategies to estimate ROPs include non-linear and closed-form approaches. Although non-linear solutions have been shown to be more robust against matching outliers [46], they are not widely used due to their requirement for good initial values for the unknowns [46]. To overcome this issue, this study exploits information from the GNSS/INS trajectory to derive initial estimates of the translation and rotation parameters relating two images, as shown in Figure 10.

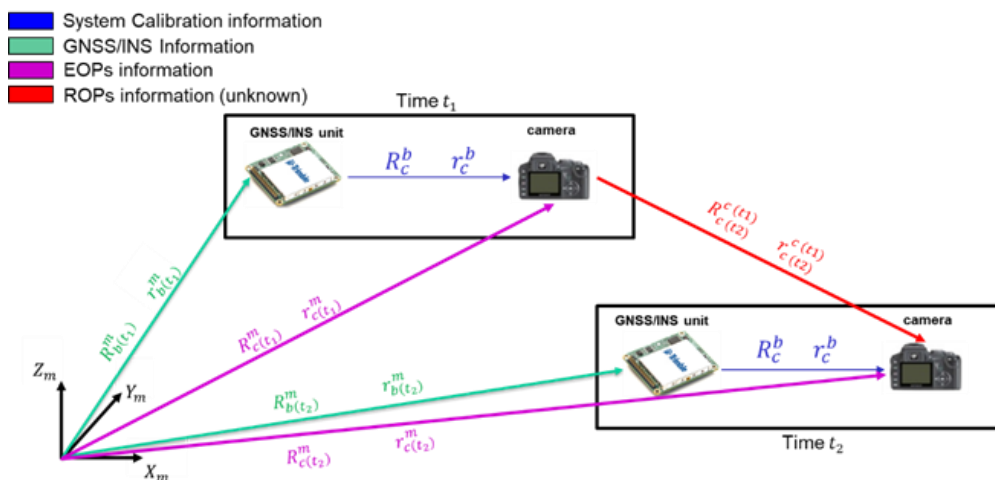


Figure 10. Illustration of relative orientation parameter (ROP) derivation from trajectory information and system calibration parameters.

To do so, we first assume that the left and right images were captured at times t_1 and t_2 , respectively. Then, image EOPs—denoted by $r_{c(t_1)}^m$ and $r_{c(t_2)}^m$ as the position vectors and $R_{c(t_1)}^m$ and $R_{c(t_2)}^m$ as rotation matrices—are derived through Equations (4a)–(4d). These EOPs are finally used to compute the ROPs between the images in question, as in Equations (5a) and (5b).

$$r_{c(t_1)}^m = r_{b(t_1)}^m + R_{b(t_1)}^m r_c^b \quad (4a)$$

$$R_{c(t_1)}^m = R_{b(t_1)}^m R_c^b \quad (4b)$$

$$r_{c(t_2)}^m = r_{b(t_2)}^m + R_{b(t_2)}^m r_c^b \quad (4c)$$

$$R_{c(t_2)}^m = R_{b(t_2)}^m R_c^b \quad (4d)$$

where:

$r_{b(t_i)}^m$ is the position of GNSS/INS body frame relative to the mapping reference frame at time t_i , as derived from the GNSS/INS integration process;

$R_{b(t_i)}^m$ is the rotation matrix from the GNSS/INS body frame to the mapping reference frame at time t_i , as derived from the GNSS/INS integration process;

r_c^b is the lever arm from the GNSS/INS body frame to camera coordinate system;

R_c^b is the rotation (boresight) matrix relating the camera to GNSS/INS body frame coordinate systems;

$r_{c(t_i)}^m$ is the position of the camera coordinate system relative to the mapping frame coordinate systems at time t_i , and

$R_{c(t_i)}^m$ is the rotation matrix from the camera frame to the mapping reference frame at time t_i .

$$r_{c(t_2)}^{c(t_1)} = \left(R_{c(t_1)}^m \right)^T \left(r_{c(t_2)}^m - r_{c(t_1)}^m \right) \quad (5a)$$

$$R_{c(t_2)}^{c(t_1)} = \left(R_{c(t_1)}^m \right)^T R_{c(t_2)}^m \quad (5b)$$

where:

$r_{c(t_2)}^{c(t_1)}$ is the translation vector between the camera coordinate systems at times t_1 and t_2 , and

$R_{c(t_2)}^{c(t_1)}$ is the rotation matrix between the camera coordinate systems at times t_1 and t_2 .

Once the approximate values for the ROPs are derived, the iterative five-point approach [58], which is based on a modified version of co-planarity constraint, is implemented. In this approach, given the initial values for the ROPs, i.e., $r_{c(t_1)}^{c(t_1)}$ and $R_{c(t_2)}^{c(t_1)}$, and assuming unknown incremental rotation and translation corrections— δR and δr , respectively—the coplanarity constraint in Equation (1), is represented by Equation (6). In Equation (6), δR is defined by the incremental angles $\Delta\omega$, $\Delta\phi$, and $\Delta\kappa$, and δr comprises the incremental translation components Δr_x , Δr_y , and Δr_z . Since for the relative orientation, the translation vector can only be determined up to an arbitrary scale, the correction to one of the translation components (say, Δr_x assuming that the baseline between the two images is mainly aligned along the x-axis of the left camera coordinate system) can be set to zero. Moreover, assuming good approximate values for the rotation angles, the incremental rotation matrix can be represented as in Equation (7). Substituting Equation (7) into Equation (6), while ignoring second-order incremental terms, results in a linear equation in five unknown parameters. Consequently, given five or more conjugate features, a least-squares solution for the unknown corrections can be derived. The derived corrections can be then used for evaluating better estimates of the ROPs, which are in turn used as approximations for another iteration to estimate another set of corrections. The iterative procedure continues until a convergence criterion is met, i.e., no significant change in the ROP estimates is observed between two successive iterations. It should be noted that having good-quality trajectory and

system calibration parameters will ensure the validity of the made assumptions (i.e., small corrections to the GNSS/INS-based ROPs). Applying the iterative five-point algorithm not only results in a refined estimation of ROPs but can also be used to simultaneously remove potential matching outliers. The outlier removal process is conducted by generating normalized image coordinates according to epipolar geometry and imposing constraints on the x-parallax and y-parallax values for the potential matches. More specifically, the minimum value for x-parallax can be set to zero owing to the fact that all object points should always lie below the camera stations. Moreover, assuming vertical imagery, the baseline-to-height ratio can be used to impose another constraint on the x-parallax values. Furthermore, conjugate points that exceed a predefined y-parallax threshold, denoted as P_y , are detected and removed as outliers. The P_y threshold depends on the quality of the GNSS/INS trajectory and system calibration parameters (i.e., smaller threshold value could be used for higher quality parameters).

$$p_1^T \cdot \left(\begin{pmatrix} r_{c(t_1)} \\ r_{c(t_2)} \end{pmatrix} + \delta r \right) \times R_{c(t_2)}^{c(t_1)} \delta R p_2 = 0 \quad (6)$$

$$\delta R = \begin{bmatrix} 1 & -\Delta\kappa & \Delta\phi \\ \Delta\kappa & 1 & -\Delta\omega \\ -\Delta\phi & \Delta\omega & 1 \end{bmatrix} \quad (7)$$

As a result of above-mentioned process, ROPs among all possible stereo-pairs are estimated and some matching outliers are removed. As mentioned earlier, in the next step of the proposed framework, two strategies are introduced, i.e., partially GNSS/INS-assisted SfM and fully GNSS/INS-assisted SfM. In the partially GNSS/INS-assisted approach, an incremental EOP recovery process based on the strategy proposed in Reference [33] is employed. This stage aims at: (1) defining EOPs in a common reference frame and (2) removing some of the matching outliers. The former is achieved through establishing a local coordinate system by selecting a stereo-pair, and then sequentially augmenting the remaining images into the final image block through rotation averaging and translation averaging strategies. The matching outlier removal is conducted within the translation averaging process where conjugate points that exhibit large back-projection errors are detected and removed. More details regarding the implemented incremental approach can be found in Reference [33].

4.3. GNSS/INS-Assisted Bundle Adjustment

In the last stage of the proposed SfM framework, a GNSS/INS-assisted bundle adjustment is conducted for parameter refinement. There are two possible scenarios, i.e., partially GNSS/INS-assisted SfM and fully GNSS/INS-assisted SfM, for conducting the bundle adjustment process, which are discussed in the following subsections.

4.3.1. Bundle Adjustment when Adopting Partially GNSS/INS-Assisted SfM

The implemented bundle adjustment procedure in this phase consists of three steps: feature tracking, 3D similarity transformation, and GNSS/INS-assisted bundle adjustment. More specifically, in the first step, SIFT matches that survived the prior steps of (1) left-to-right and right-to-left descriptor-based matching, (2) ROP estimation while removing matching outliers by imposing x-parallax and y-parallax constraints, and (3) translation averaging process while removing points with large back-projection errors, are tracked among all the involved imagery. The feature tracking process is based on a sub-graph structure, as shown in Figure 11, where the involved images are modeled as the nodes, and the estimated relative orientation parameters between the stereo-pairs represent the edges connecting different nodes. More specifically, given a point p in image i , its corresponding points are first identified within the sub-graph of image i through an exhaustive search of the derived feature correspondences in possible stereo-pairs involving image i . Then, the same procedure is sequentially conducted on every corresponding point that is identified in any of the overlapping images, i.e., images j, k, l, m, n , and q in Figure 11. This feature tracking process for point p is repeated until no

more corresponding points can be identified. In this research, an image feature has to be visible in at least three images to be considered further. Once all the SIFT features are tracked among all images, a multi light-ray intersection procedure is applied to derive the 3D coordinates of conjugate points. The outcome of this phase of data processing is a sparse point cloud, whose density depends on the feature detector and matching strategies. The coordinates of this point cloud are defined relative to the reference frame associated with the EOPs of the involved images. Before conducting bundle adjustment, a 3D similarity transformation must first be adopted to map the local frame-based EOPs to trajectory-based EOPs. Then, using the estimated transformation parameters, the sparse point cloud coordinates are transformed from the local reference frame to the trajectory-based mapping frame.

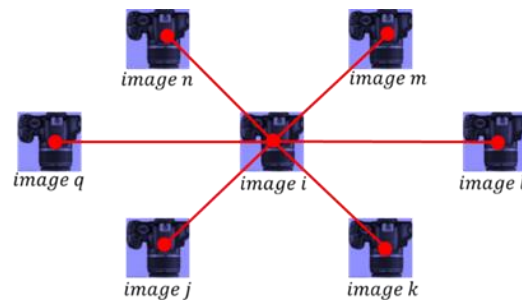


Figure 11. A sub-graph structure established on image *i*.

Finally, a GNSS/INS-assisted bundle adjustment, based on modified collinearity equations, is conducted for parameters estimation/refinement. Using the classical collinearity model presented in Equation (8), the integrated GNSS/INS position and orientation information is incorporated into the bundle adjustment procedure through a set of nonlinear equations, Equations (9a) and (9b). Equation (9a) represents the three additional equations for incorporating the GNSS/INS-based position information. The rotation matrix equality in Equation (9b) comprises nine equations, only three of which are independent due to the inherent six orthogonality constraints among the elements of a rotation matrix. Therefore, only three independent elements in this equality are used to consider the GNSS/INS-based attitude information. Thus, for a given camera at time of exposure *t*, six nonlinear equations are added to the observation equations to incorporate the GNSS/INS position and orientation information.

Unlike the classical collinearity equations, modified collinearity equations (Equation (10) and Figure 12) are used to directly incorporate the GNSS/INS position and orientation information in the bundle adjustment process.

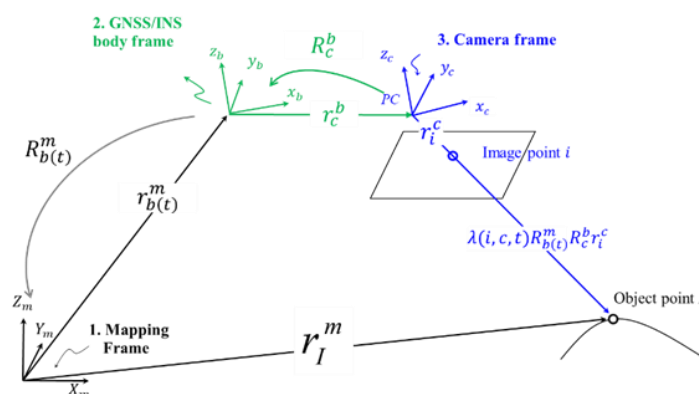


Figure 12. Illustration of the modified collinearity equations.

Using the modified collinearity equations, the available GNSS/INS position and orientation information can be added to the model through pseudo observations—i.e., direct observations of the unknowns—as presented in Equations (11a)–(11d). Comparing the additional observations in

Equations (9) and (11), it can be observed that incorporation of the available GNSS/INS information is much simpler when using the modified collinearity equations. Moreover, the modified collinearity equations can be easily extended to multi-camera systems.

$$r_I^m = r_{c(t)}^m + \lambda(i, c, t) R_{c(t)}^m r_i^c \quad (8)$$

where:

r_I^m is the ground coordinate vector of object point I ;

$r_i^c = \begin{bmatrix} x_i - x_p - dist_{x_i} \\ y_i - y_p - dist_{y_i} \\ -c \end{bmatrix}$ is the vector connecting the camera perspective center to image point i ;

x_p and y_p are the principal point coordinates of the used camera;

c is 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 , and

$\lambda(i, c, t)$ is the scale factor for point i captured by the camera at time t .

$$r_{b(t)}^m = r_{c(t)}^m + R_{c(t)}^m r_b^c + e_{r_{b(t)}^m} \quad (9a)$$

$$R_{b(t)}^m = R_{c(t)}^m R_b^c + e_{R_{b(t)}^m} \quad (9b)$$

where:

r_b^c is the lever arm from the camera coordinate system to GNSS/INS body frame;

R_b^c is the rotation (boresight) matrix from the GNSS/INS body frame to camera coordinate system;

$e_{r_{b(t)}^m}$ is the associated random error with the GNSS/INS-based position, and

$e_{R_{b(t)}^m}$ is the associated random error with the GNSS/INS-based rotation matrix, $R_{b(t)}^m$, derived through the law of error propagation using the stochastic properties of the GNSS/INS-based pitch, roll, and heading.

$$r_I^m = r_{b(t)}^m + R_{b(t)}^m r_c^b + \lambda(i, c, t) R_{b(t)}^m R_c^b r_i^c \quad (10)$$

$$r_{b(t)}^m (obs) = r_{b(t)}^m + e_{r_{b(t)}^m (obs)} \quad (11a)$$

$$Pitch_{b(t)}^m (obs) = Pitch_{b(t)}^m + e_{Pitch_{b(t)}^m (obs)} \quad (11b)$$

$$Roll_{b(t)}^m (obs) = Roll_{b(t)}^m + e_{Roll_{b(t)}^m (obs)} \quad (11c)$$

$$Heading_{b(t)}^m (obs) = Heading_{b(t)}^m + e_{Heading_{b(t)}^m (obs)} \quad (11d)$$

where:

$r_{b(t)}^m (obs)$ denotes the position observations of the GNSS/INS body frame relative to the mapping reference frame at time t , as derived from the GNSS/INS integration process;

$Pitch_{b(t)}^m (obs)$, $Roll_{b(t)}^m (obs)$, and $Heading_{b(t)}^m (obs)$ are the attitude observations of the GNSS/INS body frame relative to the mapping reference frame at time t , as derived from the GNSS/INS integration process;

$e_{r_{b(t)}^m (obs)}$, $e_{Pitch_{b(t)}^m (obs)}$, $e_{Roll_{b(t)}^m (obs)}$, and $e_{Heading_{b(t)}^m (obs)}$ are the random errors contaminating the position and orientation observations of GNSS/INS body frame, and

$r_{b(t)}^m$, $Pitch_{b(t)}^m$, $Roll_{b(t)}^m$, and $Heading_{b(t)}^m$ are the unknown position and attitude parameters of the GNSS/INS body frame relative to the mapping reference frame at time t that are solved for using the modified collinearity equations.

By establishing the modified collinearity equations for all image points as well as pseudo observations for the GNSS/INS trajectory prior information, a non-linear least-squares adjustment is conducted to refine 3D coordinates of the object points, GNSS/INS trajectory, and/or system calibration parameters. It should be noted that in cases where ground targets are available, these points are

manually identified in the images and added into the bundle adjustment process as check points for quantitative evaluations of 3D reconstruction results. Figure 13 shows the input and output for the GNSS/INS-assisted bundle adjustment process.

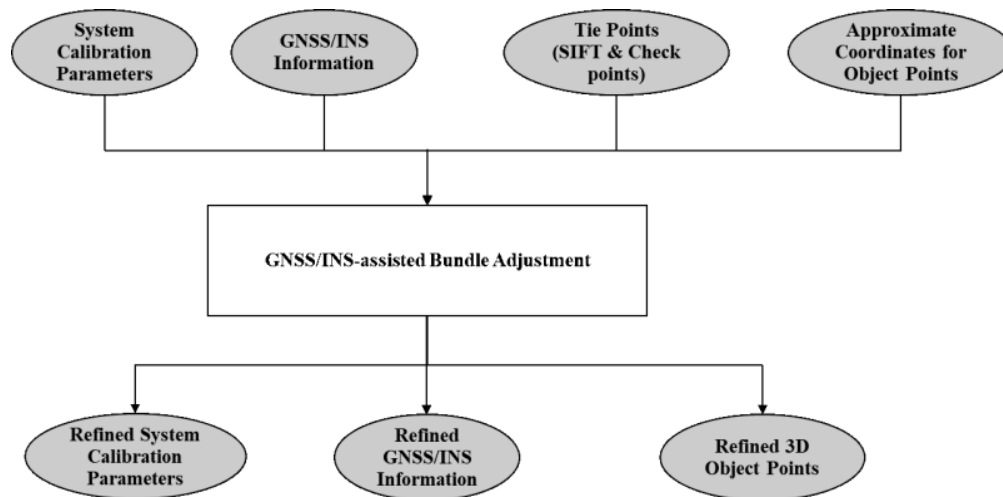


Figure 13. Input and output for the adopted GNSS/INS-assisted bundle adjustment.

4.3.2. Bundle Adjustment when Adopting Fully GNSS/INS-Assisted SfM

Although global and incremental strategies for EOP recovery perform well in the presence of accurate ROPs, they are computationally expensive, especially when dealing with datasets with a large number of images. As described in Section 4.2, the goal of the incremental EOP recovery stage is two-fold: (1) deriving the image EOPs in a local coordinate system and (2) removing matching outliers, which is conducted within the translation averaging process. With availability of GNSS/INS trajectory and system calibration parameters, image EOPs can be directly derived in the mapping frame, thus the first goal of the EOP recovery stage is already achieved. To remove the time-consuming procedure of EOP recovery while removing some matching outliers, a pre-processing RANSAC-based strategy is proposed. The proposed strategy is carried out in three steps. First, EOPs are derived using Equations (4a)–(4d), followed by a feature tracking process, as described in Section 4.3.1, to derive the corresponding features among all overlapping images. Finally, a RANSAC-based matching outlier approach is used. In the introduced approach, given a point p , which has been tracked in m images (where $m \geq 3$), two out of the m images are randomly selected and the 3D coordinates of the SIFT feature in question are derived through a two light-ray intersection procedure. Then, the normal distances, denoted as n_d , between the derived 3D point and the light-rays defined by the remaining tracked image points ($m-2$) and their respective image perspective centers are calculated. For a tracked point to be considered as an inlier for the current RANSAC trial, the distance, n_d , should be smaller than a predefined threshold, denoted as D . After repeating the above procedure for all possible trials, the sample with the largest consensus set is used together with the compatible tracked points to derive the final 3D coordinates for the SIFT feature in question. This process is shown in Figure 14. Once the matching outliers are removed and the approximate coordinates for 3D object points are derived, we proceed with the GNSS/INS-assisted bundle adjustment for parameter refinement, as described in Section 4.3.1. In the proposed strategy, the bundle adjustment is only conducted for images with more than N SIFT-based tie points. The images that have more than N tie points will be denoted hereafter as BA-input images.

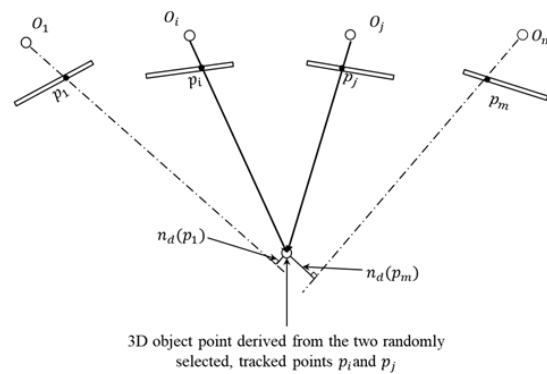


Figure 14. Proposed RANSAC procedure for final outlier removal before bundle adjustment.

5. Experimental Results and Discussion

This section evaluates the capability of the proposed SfM frameworks using eight UAV-based image datasets. In addition, the performance of the introduced SfM strategies is compared with a traditional SfM framework, which is a modified version of the approach introduced in [33]. For the modified framework, rather than conducting an indirect georeferencing, a GNSS/INS-assisted bundle adjustment is employed, as described in Section 4.3.1. C++ programming language was used for the generation of an in-house implementation of the traditional and proposed frameworks while using the SiftGPU [45] as the SIFT detector and descriptor algorithm. Table 2 describes the algorithms implemented in each step of the traditional, partially GNSS/INS-assisted, and fully GNSS/INS-assisted SfM strategies. In addition, the selected threshold values used in the traditional and proposed SfM experiments are presented in Table 3 (the same thresholds were used for all the datasets).

Table 2. Implemented algorithms in each step of the traditional and proposed SfM frameworks.

Framework	Matching Search Space	ROP Estimation	EOP Recovery	Bundle Adjustment
Traditional	Exhaustive search	Two-point + iterative five-point	Incremental strategy	GNSS/INS-assisted BA
Partially GNSS/INS-assisted	GNSS/INS-assisted, reduced search space	GNSS/INS-based ROP estimation + iterative five-point	Incremental strategy	GNSS/INS-assisted BA
Fully GNSS/INS-assisted	GNSS/INS-assisted, reduced search space	GNSS/INS-based ROP estimation + iterative five-point	N/A ¹	GNSS/INS-assisted BA

¹ Not applicable.

Table 3. Selected threshold values for the traditional and proposed SfM experiments.

Threshold Description	Threshold Value
K parameter for finding neighboring images as candidate stereo-pairs	20
d_1/d_2 threshold for SIFT feature matching	0.7
Search window size for GNSS/INS matching	500 × 500 pixels
Threshold for point-to-epipolar line distance for GNSS/INS matching	40.0 pixels
y-parallax threshold (P_y) for removing matching outliers in the iterative five-point approach	20.0 pixels
Minimum number of tracked features for an object point	3
D for RANSAC normal distance (n_d) threshold	0.2 m
N threshold for minimum number of SIFT-based tie points in an image for it to be considered in the BA process	20

In the remainder of this section, the reconstruction accuracy for ACRE-42 and ACRE-21C datasets are evaluated through check point analysis. As mentioned in Section 3, there are no checkerboard targets available for the remaining four experimental test fields (Romney, Windfall, Atlanta-1, and Atlanta-2) that can be used as check points for accuracy validation. Therefore, the acquired airborne LiDAR data is considered as a potential reference for accuracy evaluation of SfM-based reconstruction for such fields. In this regard, the reconstruction accuracy evaluation results from ACRE-42 and ACRE-21C datasets are used to assess the accuracy of the available LiDAR data by comparing the image-based and LiDAR-based point clouds. If such evaluation ascertains the accuracy of the LiDAR data, then the LiDAR-derived point clouds can be used as a reference for validating the image-based point clouds for the datasets, where no check points are available. Lastly, the performance of the proposed frameworks is also compared with Pix4D Mapper Pro using the datasets with more challenging texture conditions, i.e., Romney, Windfall, Atlanta-1, and Atlanta-2 datasets. In order to ensure a comprehensive analysis, two configurations in the “3D Maps” settings in Pix4D Mapper Pro are used: (1) “standard configuration”, which does not use any GNSS/INS information, hereafter denoted as Pix4D-1 and (2) “accurate geolocation and orientation”, which requires all images to be geolocated and oriented, hereafter denoted as Pix4D-2. When using Pix4D-2, the position and orientation of involved imagery (EOPs) are derived through Equations (4a)–(4d) and fed into the software.

5.1. SfM Results for Datasets with Check Points

In this subsection, the performance of the proposed framework is evaluated using four criteria:

- **Number of BA-input images:** This refers to the number of images surviving the preprocessing steps up to bundle adjustment. This criterion indicates the ability of the used SfM framework to successfully establish enough conjugate features among the involved imagery. More specifically, in case of incremental EOP recovery adopted in the traditional and partially GNSS/INS-assisted SfM approaches, this number indicates all the images for which the EOPs have been successfully estimated. On the other hand, for the fully GNSS/INS-assisted SfM framework, this criterion pertains to the number of images with more than N SIFT-based tie points (where $N > 20$).
- **Number of reconstructed object points:** This criterion represents the sparsity/density of the SfM-based point cloud.
- **Square root of a-posteriori variance factor resulting from the bundle adjustment process ($\hat{\sigma}_0$):** This value illustrates the quality of fit between observations and unknowns as represented by the mathematical model of the modified collinearity equations. Small values for $\hat{\sigma}_0$ serve as an indication of small image residuals as a result of the back-projection process using the estimated unknowns.
- **$X/Y/Z_{RMSE}$:** The root-mean-square error (RMSE) values of differences between the bundle adjustment-derived and surveyed coordinates of the check points reflect the accuracy of the 3D reconstruction.

The above-mentioned evaluation criteria pertaining to the traditional, partially GNSS/INS-assisted and fully GNSS/INS-assisted SfM techniques as well as the total number of acquired images for the four ACRE-42 and ACRE-21C datasets are presented in Table 4. As can be seen in this table, the traditional SfM fails to estimate EOPs for 12 to 45 images, i.e., 5% to 9% of the total images, while both the proposed frameworks are able to incorporate all the captured images in all experiments. Inspecting the number of object points in Table 4, it can be observed that the proposed frameworks result in point clouds with more points when compared to the traditional SfM for all the datasets, with the fully GNSS/INS-assisted producing the largest set of 3D points. Figure 15 illustrates two sample images together with reconstructed object points and successfully processed image locations resulting from the traditional and fully GNSS/INS-assisted approaches for the August dataset over the ACRE-42 test field (since the two proposed frameworks show similar performance, only results from the fully GNSS/INS-assisted framework are presented in this figure). As depicted in Figure 15a,b, the failure

of the traditional SfM to recover EOPs for some images, which in turn could induce gaps in the final 3D point clouds and orthophotos, takes place at regions that cover densely planted plots with similar genotypes (right image in Figure 15b). Figure 15c shows the ability of the fully GNSS/INS-assisted framework in processing all the involved imagery and generating 3D object points, even in parts of the test field with challenging texture patterns.

Table 4. Performance comparison of the traditional and proposed SfM frameworks in terms of the number of BA-input images, number of derived object points, square root of a-posteriori variance factor, $\hat{\sigma}_0$, and root-mean-square error (RMSE) of differences between the BA-based and surveyed coordinates of the check points for ACRE-42 and ACRE-21C datasets.

Dataset	Total # of Images	SfM Technique	# of BA-Input Images	# of Object Points	$\hat{\sigma}_0$ (pixel)	X_{RMSE} (m)	Y_{RMSE} (m)	Z_{RMSE} (m)
ACRE-42 20190810	562	Traditional	517	107,000	1.63	0.03	0.02	0.04
		Partially ¹	562	214,000	2.31	0.02	0.03	0.03
		Fully ²	562	352,000	4.78	0.03	0.05	0.03
ACRE-42 20190905	555	Traditional	525	123,000	1.56	0.04	0.03	0.04
		Partially	555	331,000	2.47	0.04	0.03	0.02
		Fully	555	464,000	4.47	0.04	0.03	0.03
ACRE-21C 20190823	215	Traditional	202	55,000	1.77	0.02	0.02	0.03
		Partially	215	101,000	2.36	0.02	0.02	0.02
		Fully	215	175,000	4.97	0.02	0.03	0.03
ACRE-21C 20190904	217	Traditional	198	74,000	2.01	0.03	0.01	0.03
		Partially	217	149,000	2.99	0.03	0.02	0.03
		Fully	217	210,000	4.87	0.03	0.01	0.05

¹ Partially GNSS/INS-assisted SfM. ² Fully GNSS/INS-assisted SfM.

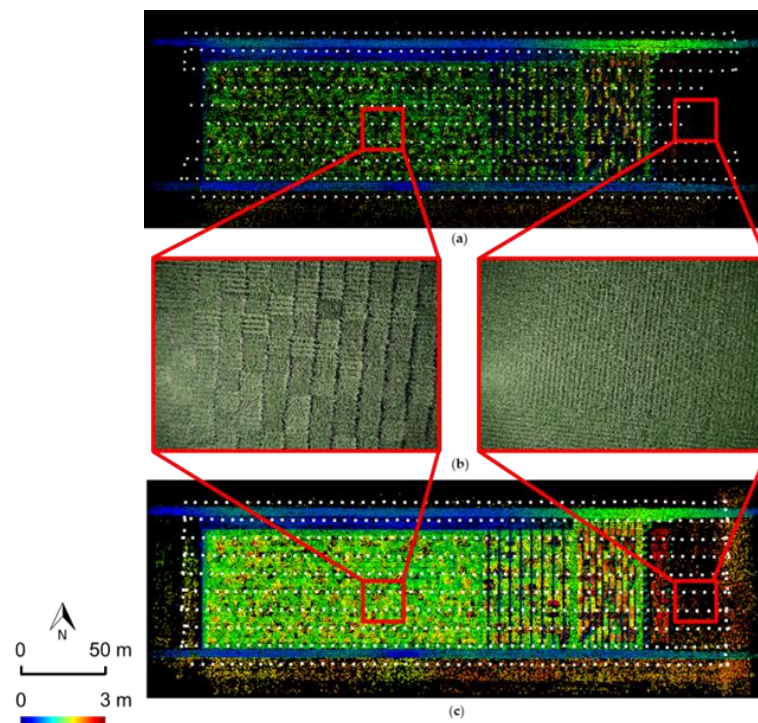


Figure 15. Top view of the reconstructed object points (colored by height) and the successfully processed image locations (white dots) for August dataset over ACRE-42: (a) the traditional SfM with 107,000 points, (b) sample images with different texture patterns, and (c) the fully GNSS/INS-assisted SfM with 352,000 points.

In terms of the derived $\hat{\sigma}_0$ values, it can be noted that these values for all the experiments are smaller than 5.0 pixels, i.e., 0.04 m in the object space. More specifically, the traditional SfM achieves the best value for $\hat{\sigma}_0$ with a range of 1.56 to 2.01 pixels, while the fully GNSS/INS-assisted SfM produces the largest $\hat{\sigma}_0$ values with a range of 4.47 to 4.97 pixels. Owing to the fact that the proposed approaches lead to a greater number of BA-input images and object points, the larger image residuals—which are key contributor to the $\hat{\sigma}_0$ value estimation—in these frameworks can be considered trivial. This fact is also supported by the reconstruction accuracy results reported in Table 4. The check point analysis indicates that planimetric and vertical RMSE values are smaller than 0.05 m, thus indicating a good accuracy of the derived 3D object points from all the implemented SfM frameworks. Therefore, it can be concluded that the three approaches exhibit similar performance for all the datasets (i.e., they exhibit similar accuracy at the location of check points).

In addition to the above-mentioned analyses, a comparison based on the processing time is conducted among the three approaches. Table 5 reports the individual processing time for the three frameworks for each of the following steps, namely:

- Image matching and ROP estimation: This process includes the SIFT detector and descriptor evaluation, feature matching, and ROP estimation while removing some matching outliers. One should note that the partially GNSS/INS-assisted and fully GNSS/INS-assisted frameworks share the same implementations for those steps. Therefore, both frameworks will have the same time consumption.
- BA preparation: When using the partially GNSS/INS-assisted and traditional frameworks, the BA preparation pertains to the incremental approach for EOP recovery, feature tracking, and 3D similarity transformation to bring the derived 3D points from the local frame to the GNSS/INS trajectory reference frame. Given that the traditional and partially GNSS/INS-assisted frameworks deal with a different number of images and object points, the time consumption for these approaches is not expected to be identical. For fully GNSS/INS-assisted SfM, the BA preparation refers to the feature tracking and RANSAC procedure for removing matching outliers.
- BA: This process refers to the implemented GNSS/INS-assisted bundle adjustment process.

Table 5. SfM processing time for the ACRE-42 and ACRE-21C datasets.

Dataset	SfM Technique	Processing Time (min)			
		Image Matching and ROP Estimation	BA Preprocessing	BA	Total
ACRE-42 20190810	Traditional	42.2	16.9	6.1	65.2
	Partially	56.1	50.2	11.1	117.4
	Fully	56.1	14.5	18.8	89.4
ACRE-42 20190905	Traditional	42.2	24.7	12.4	79.3
	Partially	71.6	52.3	21.5	145.4
	Fully	71.6	17.0	31.4	120.0
ACRE-21C 20190823	Traditional	14.4	3.1	3.3	20.8
	Partially	26.9	13.6	5.3	45.8
	Fully	26.9	6.2	11.1	44.2
ACRE-21C 20190904	Traditional	25.0	4.4	4.2	33.6
	Partially	28.5	14.7	7.4	50.6
	Fully	28.5	7.5	11.1	47.1

As shown in Table 5, the traditional framework is faster than the proposed approaches in executing the image matching and ROP estimation steps. This performance is mainly due to the fact that the proposed image matching and ROP estimation strategies result in more conjugate points and stereo-pairs. For example, while the traditional SfM establishes 794 stereo-pairs with an average of 618

conjugate features per stereo-pair for the August dataset over the ACRE-21C test field, the proposed image matching and ROP estimation strategies result in a total of 2178 stereo-pairs with an average of 1115 conjugate points per stereo-pair for the same dataset. Another possible explanation for longer matching processing time of the proposed frameworks could be the fact that the introduced GNSS/INS-assisted matching strategy conducts the forward-backward projections for each extracted SIFT feature. This can be a time-consuming process when dealing with a large set of features.

As hypothesized earlier, even though the same implementation of BA preparation has been used for the traditional and partially GNSS/INS-assisted frameworks, the former is observed to execute in a shorter time as it deals with a smaller set of conjugate points and stereo-pairs. The processing times reported in Table 5 demonstrate that the fully GNSS/INS-assisted SfM is more efficient than its partially GNSS/INS-assisted counterpart in terms of the BA preparation step, even though the former is dealing with a larger number of 3D points. This can be explained by the fact that the fully GNSS/INS-assisted SfM eliminates the requirement for the significantly time-consuming EOP recovery step by exploiting the available GNSS/INS trajectory. The last component—bundle adjustment—is found to be more computationally intensive for the proposed methods, with the longest execution time corresponding to the fully GNSS/INS-assisted framework. This can be attributed to the increased number of BA-input images and SIFT-based tie points for the proposed frameworks, thus leading to a larger number of observation equations in the bundle adjustment procedure. For example, the number of reconstructed SIFT-based tie points from the traditional, partially GNSS/INS-assisted and fully GNSS/INS-assisted approaches for the ACRE-42 August dataset was 381,000, 797,000, and 1,348,000, respectively. However, considering the number of reconstructed points, the BA processing time for these approaches was 0.16, 0.14, and 0.14 min/10,000 tie points, respectively. Such results show similar performance of the three SfM frameworks in terms of the BA execution time when considering the number of SIFT-based tie points. Overall, from Tables 4 and 5, one can conclude that without a significant increase in the processing time, the proposed frameworks can lead to more favorable reconstruction results, namely, generating accurate 3D point clouds with larger size and incorporating the entire captured images in the SfM procedure. Having more images and object points is critical for the generation of an orthophoto with high quality. Also, comparing the two proposed approaches, the fully GNSS/INS-assisted SfM framework has been shown to be more efficient than the partially GNSS/INS-assisted approach, especially when dealing with datasets including a large number of images.

As mentioned earlier in this section, this study also aims at investigating whether the available LiDAR data—acquired simultaneously with the images—could be used as a potential reference for accuracy evaluation of image-based point clouds. If confirmed, using the LiDAR data as a reference will be valuable in the absence of previously established check points. Moreover, using the LiDAR data as a reference is more advantageous as it allows for evaluating the reconstruction quality over the entire test field instead of only at the check point locations. In this regard, using the accuracy evaluation results from the datasets with checkerboard targets, ACRE-42 and ACRE-21C, the quality of the LiDAR data is evaluated through comparison with the image-based point clouds. One should note that only a height comparison is carried out in this study since it is not possible to establish point-to-point correspondences between the LiDAR and image point clouds due to the fact that there is no significant height variation in the experimental fields. The LiDAR point cloud is first used to generate a digital surface model (DSM) with a resolution chosen according to the LiDAR point density (4 cm in this analysis). To mitigate the impact of noisy points in the LiDAR data, the DSM is generated using the 90th percentile height within each grid cell instead of the maximum height. Next, the Z coordinate of each point in the image-based point cloud is compared with its corresponding grid height in the LiDAR-based DSM. In order to ensure that the correspondence of the image-based and LiDAR-based 3D points are being compared, the height comparison is conducted only when the difference between the two Z coordinates is smaller than a predefined threshold. A Z-difference larger than the threshold is hypothesized to be either a result of noisy points or points belonging to different surfaces, such as one corresponding to a crop and another corresponding to the underlying ground.

In this study, the threshold value is selected as 20 cm while considering the accuracy of the GNSS/INS trajectory, system calibration parameters, and LiDAR measurements.

Although the above-mentioned strategy removes some of the wrong matches between the LiDAR-based DSM and image-based point cloud, the Z-difference comparison is still susceptible to artifacts arising from the fact that LiDAR- and SfM-based points are not representing the same features. Therefore, one should expect that the evaluated Z-differences might lead to pessimistic estimates—including casual biases—of the discrepancies, which should be considered while inspecting the reported statistical values. The mean, standard deviation (STD), and RMSE of differences between the Z coordinates of the SfM-based point cloud and the LiDAR DSM are presented in Table 6, where an average of 80% of the former point clouds have been matched with the latter for the four datasets. The mean values of the Z differences presented in Table 6 show that the SfM-based object points are compatible with the LiDAR point cloud. Also, the RMSE values with a range of 0.08 to 0.10 m can be attributed to the error budget of LiDAR data (i.e., ± 10 cm, considering 45 m flying height) as well as image-based reconstruction. The accuracy assessment results in Tables 4 and 6 serve as sufficient evidence to claim that the acquired LiDAR point cloud is accurate enough to be used as an elevation reference for further datasets that do not have check points.

Table 6. Mean, standard deviation (STD), and RMSE of the differences between image-based 3D points and their corresponding LiDAR-based digital surface model (DSM) grids.

Dataset	SfM Technique	Z_{MEAN} (m)	Z_{STD} (m)	Z_{RMSE} (m)
ACRE-42 20190810	Traditional	−0.06	0.07	0.09
	Partially	−0.03	0.09	0.09
	Fully	−0.02	0.10	0.10
ACRE-42 20190905	Traditional	−0.06	0.06	0.08
	Partially	−0.03	0.08	0.09
	Fully	−0.03	0.09	0.09
ACRE-21C 20190823	Traditional	−0.04	0.07	0.08
	Partially	−0.02	0.08	0.08
	Fully	−0.02	0.09	0.09
ACRE-21C 20190904	Traditional	−0.04	0.08	0.09
	Partially	−0.02	0.09	0.09
	Fully	−0.03	0.10	0.10

In addition to the above height comparison, an alignment check was conducted between a LiDAR-based DSM and the corresponding orthophoto in order to evaluate the planimetric agreement between the LiDAR and image-based reconstruction. Just as an example, Figure 16 shows a horizontal alignment check for the Romney dataset, where the XY-shifts between the LiDAR-based DSM and orthophoto are manually measured at the four corners of the experimental field (i.e., the shifts in the XY-directions between the edges of the field in the LiDAR-based DSM and corresponding orthophoto are used to evaluate the planimetric agreement between these products). Based on the reported values in Figure 16, the RMSE estimates of the planimetric differences between the two data sources are 0.04 m and 0.06 m in the X and Y directions, respectively. Similar results hold true for the remaining datasets. Therefore, one can conclude that there is a good 3D agreement between the RGB-based reconstruction and LiDAR point cloud.

5.2. SfM Results for Datasets without Check Points

In this section, the experimental results for Romney, Windfall, Atlanta-1, and Atlanta-2 datasets, where no checkerboard targets are available, are presented. Having verified the accuracy of the acquired LiDAR data in Section 5.1, the LiDAR-based validation process is carried out for elevation accuracy evaluation of SfM-based object points. Table 7 presents the comparative results of the traditional,

partially GNSS/INS-assisted, and fully GNSS/INS-assisted SfM frameworks in terms of the number of reconstructed object points and square root of a-posteriori variance factor, $\hat{\sigma}_0$, together with the mean, STD, and RMSE of Z differences between the coordinates of the image-based point cloud and the LiDAR DSM. The number of successfully incorporated images in the SfM process (BA-input images) are discussed in the next subsection where a comparison between the implemented SfM frameworks and Pix4D Mapper Pro is presented.

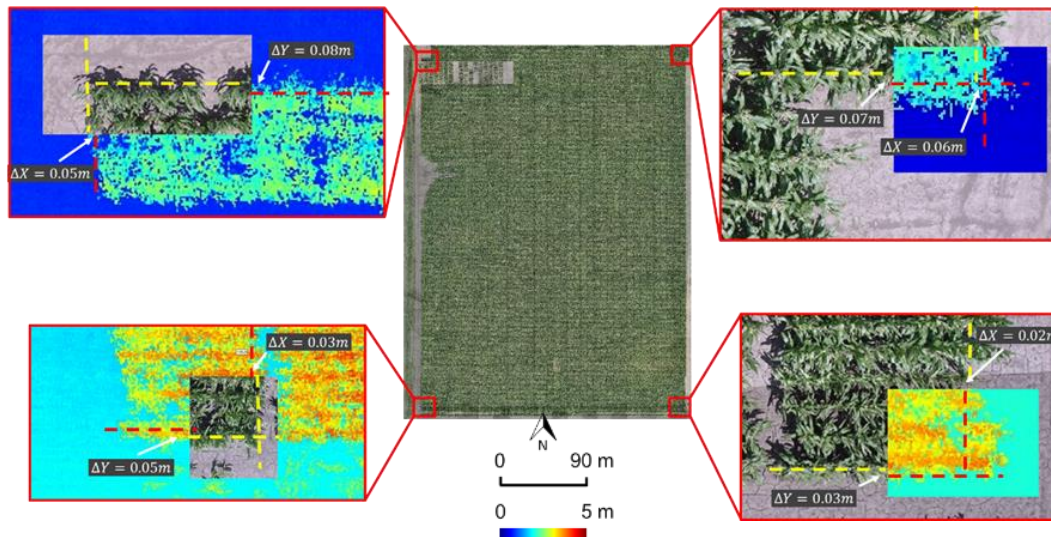


Figure 16. Horizontal alignment assessment between the LiDAR-based DSM (colored by height) and orthophoto generated from the fully GNSS/INS-assisted SfM for the Romney dataset.

Table 7. Performance comparison of the traditional and proposed SfM frameworks in terms of the number of derived object points, square root of a-posteriori variance factor, $\hat{\sigma}_0$, and mean, STD, and RMSE of the Z differences between the image-based point cloud and LiDAR-based DSM for Romney, Windfall, Atlanta-1, and Atlanta-2 datasets.

Dataset	SfM Technique	Total # of Images	# of Object Points	$\hat{\sigma}_0$ (pixel)	Z_{MEAN} (m)	Z_{STD} (m)	Z_{RMSE} (m)
Romney	Traditional		53,000	1.81	−0.01	0.07	0.08
	Partially	846	428,000	2.80	−0.01	0.12	0.12
	Fully		731,000	4.31	0.01	0.10	0.10
Windfall	Traditional		39,000	1.69	0.01	0.11	0.11
	Partially	441	286,000	3.02	0.01	0.09	0.10
	Fully		351,000	4.65	0.01	0.10	0.10
Atlanta-1	Traditional		33,000	1.60	−0.01	0.07	0.08
	Partially	254	93,000	2.04	−0.05	0.08	0.09
	Fully		152,000	4.25	−0.04	0.09	0.09
Atlanta-2	Traditional		125,000	1.00	0.02	0.08	0.09
	Partially	355	261,000	1.89	−0.04	0.07	0.08
	Fully		300,000	3.07	−0.04	0.08	0.09

As can be seen in Table 7, the fully GNSS/INS-assisted SfM shows superior performance in terms of the number of reconstructed object points, which is in keeping with the same claim being made about this strategy for the previous datasets. In terms of the $\hat{\sigma}_0$ values reported in Table 7, one can note that similar to the results presented in Section 5.1, all approaches lead to $\hat{\sigma}_0$ values smaller than 5 pixels, which is an indication of small back-projection errors (less than 0.04 m) for all datasets. In addition, based on a Z-difference threshold of 0.20 m, an average of 80% of the points belonging to the SfM-based point are matched with their corresponding 4 cm LiDAR-based DSM grid. These matches contribute

towards the height accuracy evaluation statistics reported in Table 7. The small values for the mean Z difference (−5 to +2 cm) for all the datasets when using either of the three implemented SfM frameworks indicate a good alignment between the image-based point clouds and LiDAR data. Comparing the mean Z differences reported in Tables 6 and 7, it can be seen that the small bias observed for the ACRE-42 and ACRE-21C datasets is not present for the rest of the datasets. This is mainly due to the fact that in datasets with higher plant density, i.e., Romney, Windfall, Atlanta-1, and Atlanta-2, most of LiDAR points belong to the top of the plants and this can reduce the derived shift between the LiDAR point cloud (and consequently the LiDAR-based DSM) and image-based point cloud. Also, looking into the RMSE of the Z differences, one can note that these values are within the noise level of the LiDAR data. Therefore, we conclude that the three SfM approaches lead to equally accurate 3D reconstruction results for those datasets. It should be noted that this evaluation is only valid where an object space is reconstructed from imagery. In regions with gaps, no claim can be made regarding the reconstruction quality.

Figures 17–20 present the image-based point clouds reconstructed using the traditional and proposed SfM approaches as well as the corresponding LiDAR point clouds for the Romney, Windfall, Atlanta-1, and Atlanta-2 datasets. As can be seen in these figures, while the traditional approach leads to an adequate number of object points at the perimeter of the field, where distinctive features are prevalent, it fails to reconstruct enough points within the test field (see Figures 17a, 18a, 19a and 20a). This limitation is addressed to some degree by the proposed partially GNSS/INS-assisted framework (see Figures 17b, 18b, 19b and 20b) and to a higher degree by the proposed fully GNSS/INS-assisted framework (see Figures 17c, 18c, 19c and 20c). A closer inspection reveals that the point density of the image-based point clouds can differ throughout the test field depending on the planting density, size of the plot that has a specific genotype, and similarity between genotypes among neighboring plots. This trend is more prevalent in the reconstructed point cloud for the Atlanta-2 dataset (containing two different crop types, soybean and maize) using the traditional SfM strategy (see Figure 20a). More specifically, since the soybean is planted sparsely over the field, the traditional approach is able to reconstruct a decent number of object points in this part of the test field. However, the part of the field planted with maize exhibits a repetitive pattern due to the high planting density, thus inducing ambiguities in the image matching process.

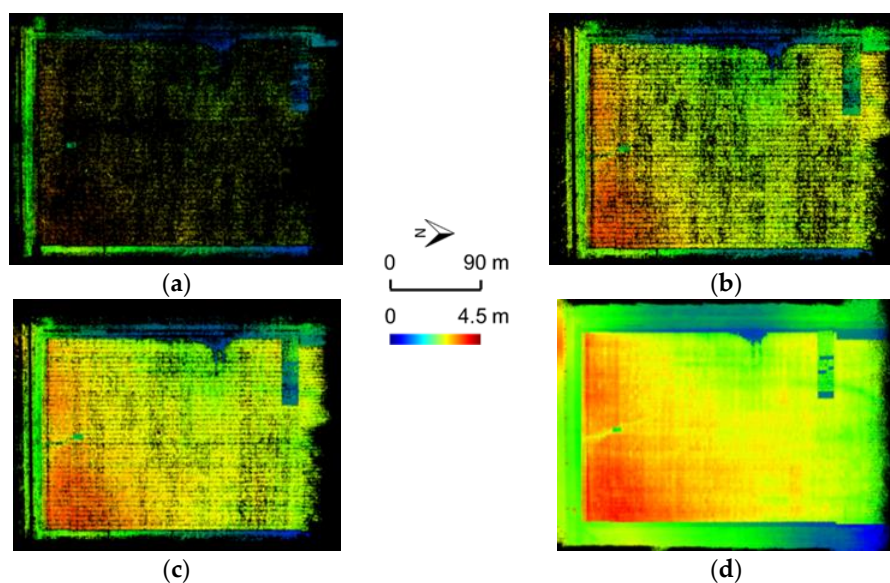


Figure 17. Image-based and LiDAR point clouds for Romney dataset with an approximate area of 100,000 m², (a) traditional SfM with 53,000 points, (b) partially GNSS/INS-assisted SfM with 428,000 points, (c) fully GNSS/INS-assisted SfM with 731,000 points, and (d) LiDAR point cloud with 134,000,000 points.

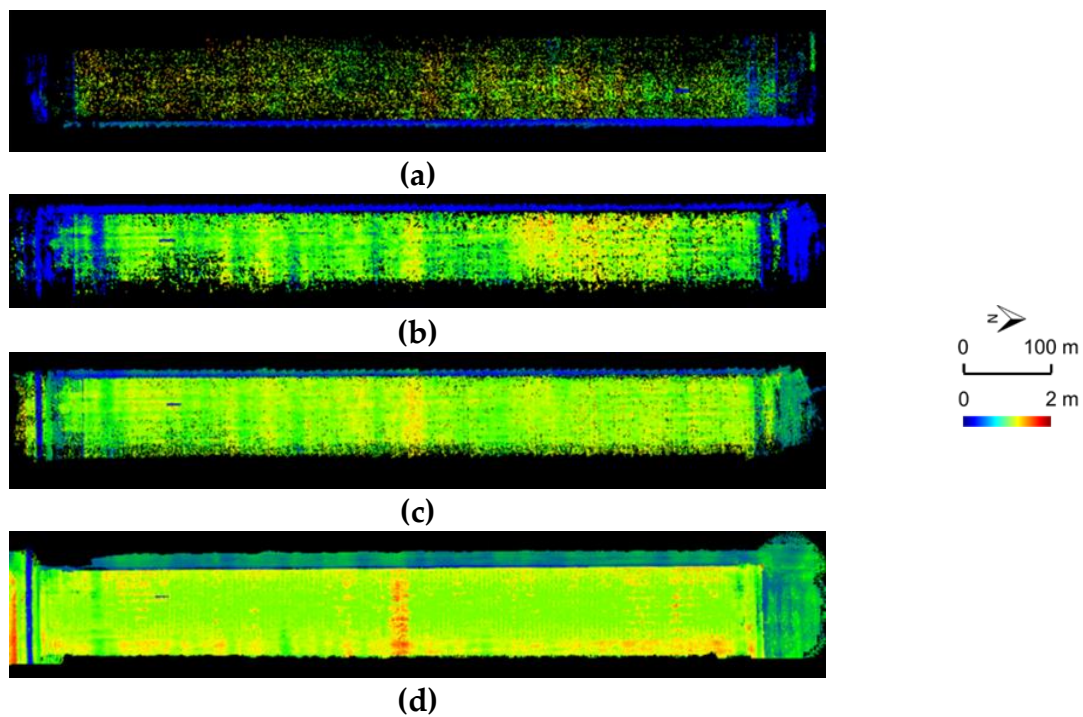


Figure 18. Image-based and LiDAR point clouds for Windfall dataset with an approximate area of $70,000 \text{ m}^2$, (a) traditional SfM with 39,000 points, (b) partially GNSS/INS-assisted SfM with 286,000 points, (c) fully GNSS/INS-assisted SfM with 351,000 points, and (d) LiDAR point cloud with 189,000,000 points.

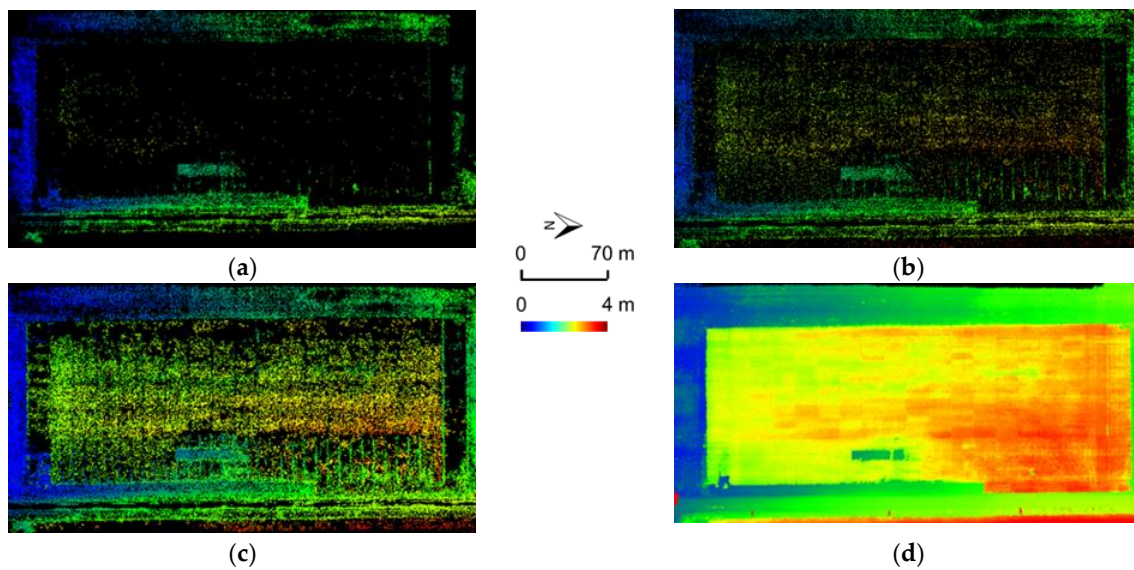


Figure 19. Image-based and LiDAR point clouds for Atlanta-1 dataset with an approximate area of $26,000 \text{ m}^2$, (a) traditional SfM with 33,000 points, (b) partially GNSS/INS-assisted SfM with 93,000 points, (c) fully GNSS/INS-assisted SfM with 152,000 points, and (d) LiDAR point cloud with 37,000,000 points.

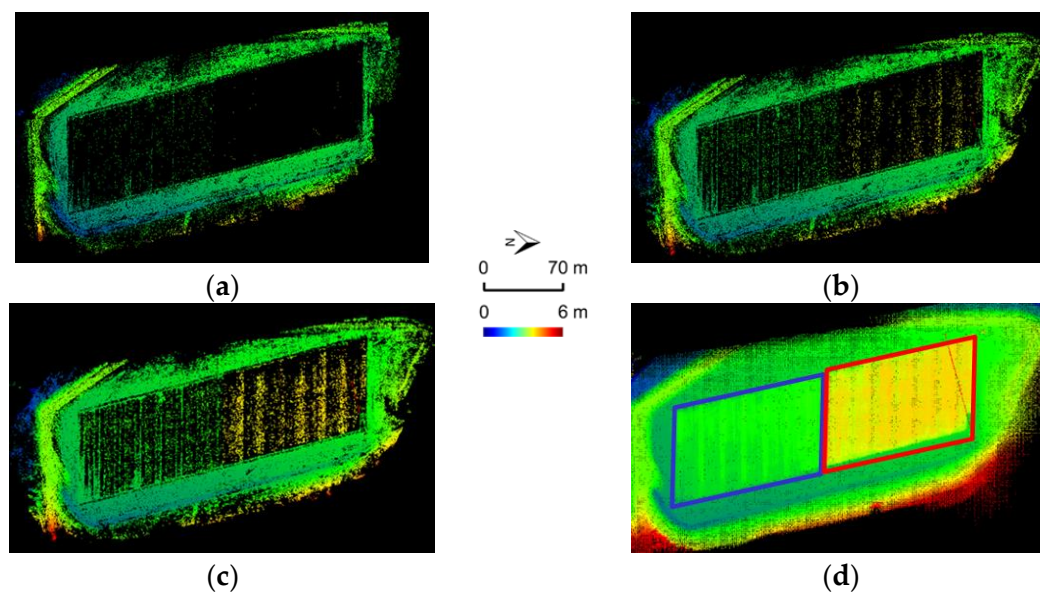


Figure 20. Image-based and LiDAR point clouds for Atlanta-2 dataset with an approximate area of 35,000 m², (a) traditional SfM with 125,000 points, (b) partially GNSS/INS-assisted SfM with 261,000 points, (c) fully GNSS/INS-assisted SfM with 300,000 points, and (d) LiDAR point cloud with 52,000,000 points: superimposed blue and red parallelograms show the soybean and maize plots, respectively.

5.3. Comparison with Pix4D Mapper Pro

In this phase of the experimental results, we compare the performance of the proposed frameworks with the traditional SfM approach, Pix4D-1 (with “standard configuration”), and Pix4D-2 (with “accurate geolocation and orientation”). The datasets used for conducting these comparative studies include Romney, Windfall, Atlanta-1, and Atlanta-2 datasets. The quantitative evaluation is conducted based on the number of BA-input images, while the qualitative analysis is performed by visual assessment of the generated orthophotos. Table 8 shows the number of BA-input images for the four datasets using the traditional SfM strategy, two proposed SfM frameworks, Pix4D-1, and Pix4D-2. As reported in Table 8, the frameworks that are not aided by the GNSS/INS information fail to process a considerable number of images for all the datasets. This trend is very evident especially for the Romney dataset, where 126 (15%) and 552 (65%) image EOPs are not recovered using traditional SfM and Pix4D-1, respectively. On the other hand, GNSS/INS-assisted frameworks, i.e., partially GNSS/INS-assisted, fully GNSS/INS-assisted, and Pix4D-2, produce better results. Amongst the three GNSS/INS-assisted frameworks, the two strategies proposed in this research show a superior performance over Pix4D-2. More specifically, while the proposed frameworks result in all the images surviving the preprocessing steps up to the bundle adjustment stage, Pix4D-2 fails to achieve the same outcome and it leaves out 4, 5, and 62 images for the Atlanta-1, Atlanta-2, and Romney datasets, respectively. The observed inferior performance of traditional SfM, Pix4D-1, and Pix4D-2 for the Romney dataset can be explained by the presence of repetitive texture patterns caused by the high planting density as well as very similar genotypes being planted along the rows for this test field. As far as the number of derived object points, Pix4D-1 and Pix4D-2 produced a similar number of points. This indicates that prior information regarding the image EOPs within Pix4D-2 is not used to identify more matches among overlapping images. Pix4D-2 reconstructs 173,000, 118,000, 223,000, and 1,000,000 object points for Romney, Windfall, Atlanta-1, and Atlanta-2 datasets, respectively. The distribution of the Pix4D-2-based object points for the different test fields is shown in Figure 21. Comparing these numbers with the quantity of reconstructed points in Table 7, it can be observed that for two datasets, i.e., Atlanta-1 and Atlanta-2, Pix4D-2 outperforms the two proposed frameworks. However, looking into the point clouds generated by Pix4D-2 in Figure 21, it can be observed that the majority of the reconstructed points belong to the area at the perimeter of the experimental fields exhibiting rich texture conditions.

The difference between these approaches can be also seen in Figure 22, where point density maps with a grid size of 0.5 m are generated for the derived point clouds from the fully GNSS/INS-assisted SfM and Pix4D-2 for the Romney dataset. In this figure, it is quite clear that the proposed fully GNSS/INS-assisted SfM produces higher point density throughout the field. In summary, inspecting Figures 17–22, one can conclude that when compared to Pix4D-2 results, the proposed frameworks result in more comprehensive and better-distributed point clouds for agricultural test fields.

Table 8. Number of BA-input images for Romney, Windfall, Atlanta-1, and Atlanta-2 datasets using the traditional SfM, partially GNSS/INS-assisted SfM, fully GNSS/INS-assisted SfM, Pix4D-1, and Pix4D-2.

Dataset	Total # of Images	Number of BA-Input Images				
		Traditional	Partially	Fully	Pix4D-1	Pix4D-2
Romney	846	720	846	846	294	784
Windfall	441	409	441	441	263	441
Atlanta-1	254	196	254	254	173	251
Atlanta-2	355	316	355	355	328	350

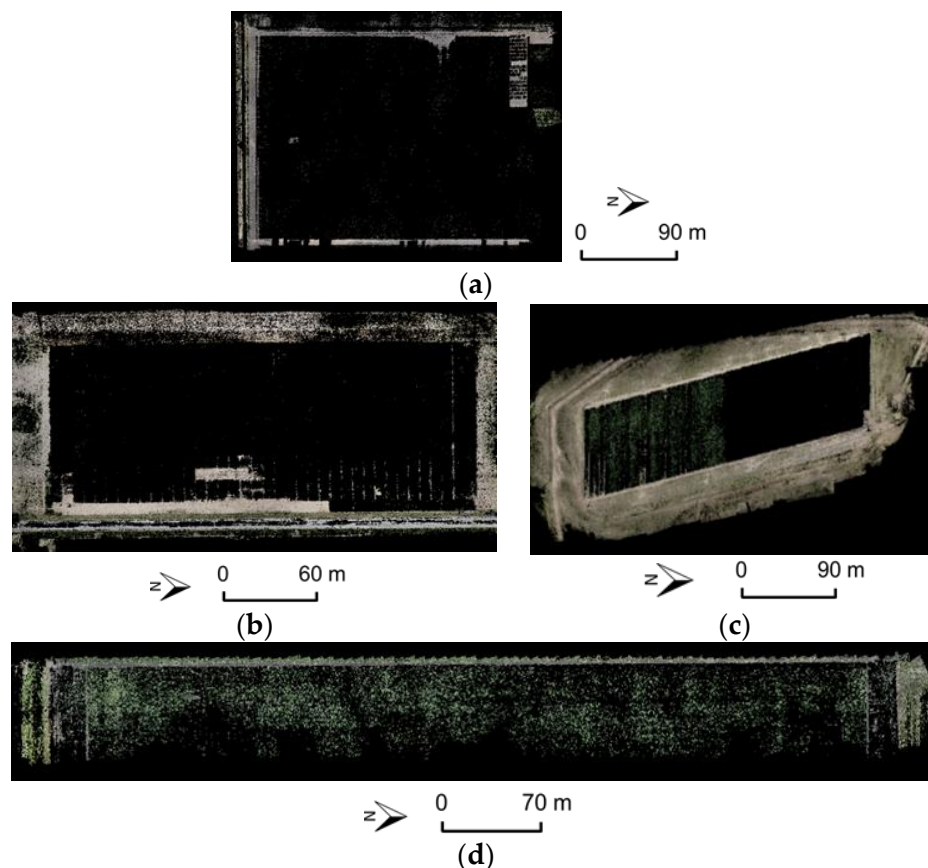


Figure 21. RGB color coded point clouds from Pix4D-2 results for (a) Romney, (b) Atlanta-1, (c) Atlanta-2, and (d) Windfall.

As mentioned earlier, the qualitative evaluation of the different strategies is done by visually assessing the orthophotos generated using the results obtained from each of the strategies under consideration—the traditional SfM, two proposed GNSS/INS-assisted frameworks, Pix4D-1, and Pix4D-2. One should note that the orthophotos generated using the results from the two proposed GNSS/INS-assisted frameworks exhibit very similar quality and so, only the orthophotos resulting from the fully GNSS/INS-assisted SfM are presented in this section. Figures 23–26 illustrate the orthophotos for the four experimental datasets. It is worth noting that since no color balancing is employed in the

proposed frameworks, obvious boundaries among the set of mosaicked images can be observed in the derived orthophotos. Obvious gaps can be observed in Figures 23–26 within all the orthophotos that are generated using frameworks which do not utilize the GNSS/INS trajectory. As already mentioned before, these gaps are mainly caused by the failure of the traditional techniques in recovering the EOPs for a considerable number of images. Looking into orthophotos generated using Pix4D-2, one can note that although Pix4D-2 was unable to recover the EOPs for all the involved images for the Atlanta-1 and Atlanta-2 datasets, no gaps can be observed in those orthophotos. This can be explained by the high percentage of overlap and side-lap between the images. Also, while the fully GNSS/INS-assisted SfM results in orthophotos without any gaps for all the experimental datasets, the generated orthophotos for the Romney and Windfall datasets using Pix4D-2 exhibit obvious gaps, especially for the Romney dataset, where Pix4D-2 was not able to recover the EOPs for 62 images.

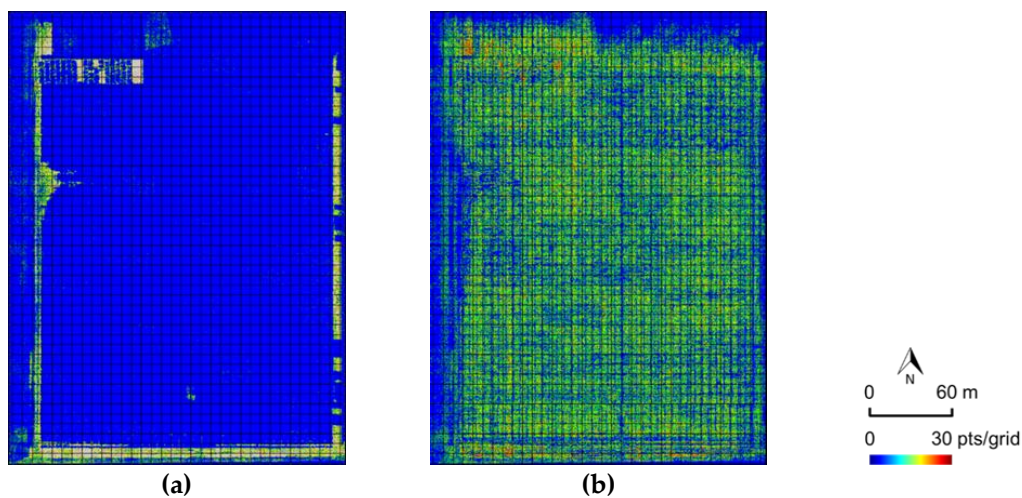


Figure 22. Point density maps generated from (a) Pix4D-2 and (b) fully GNSS/INS-assisted SfM for Romney dataset.

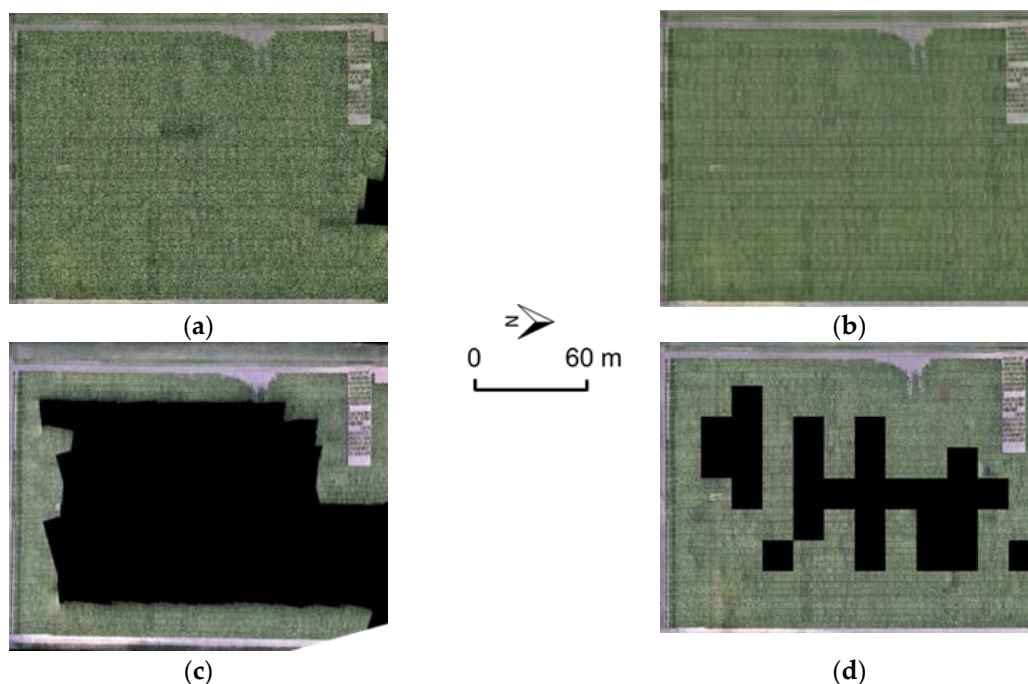


Figure 23. Romney orthophotos generated through (a) traditional SfM, (b) fully GNSS/INS-assisted SfM, (c) Pix4D-1, and (d) Pix4D-2.

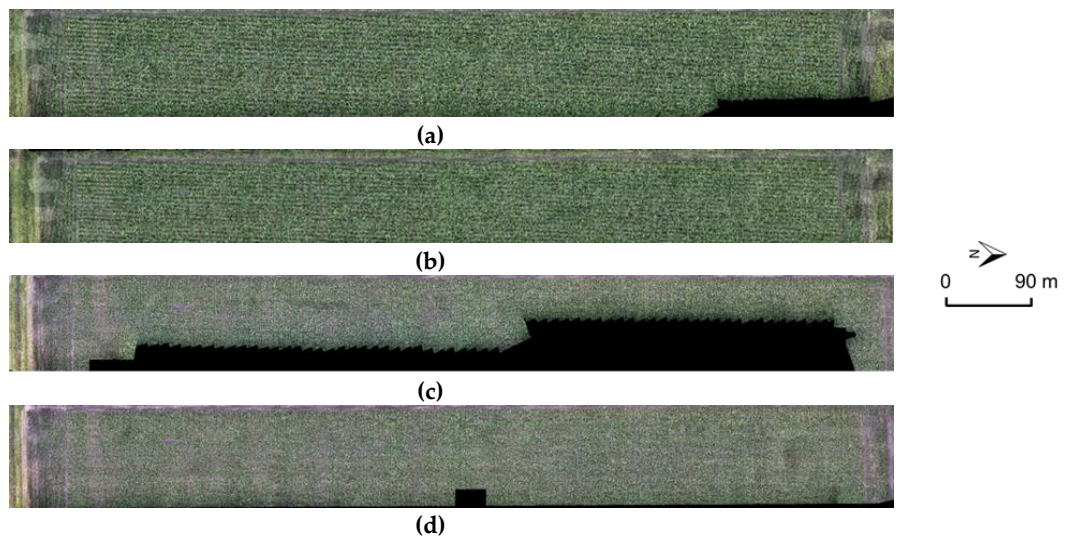


Figure 24. Windfall orthophotos generated through (a) traditional SfM, (b) fully GNSS/INS-assisted SfM, (c) Pix4D-1, and (d) Pix4D-2.

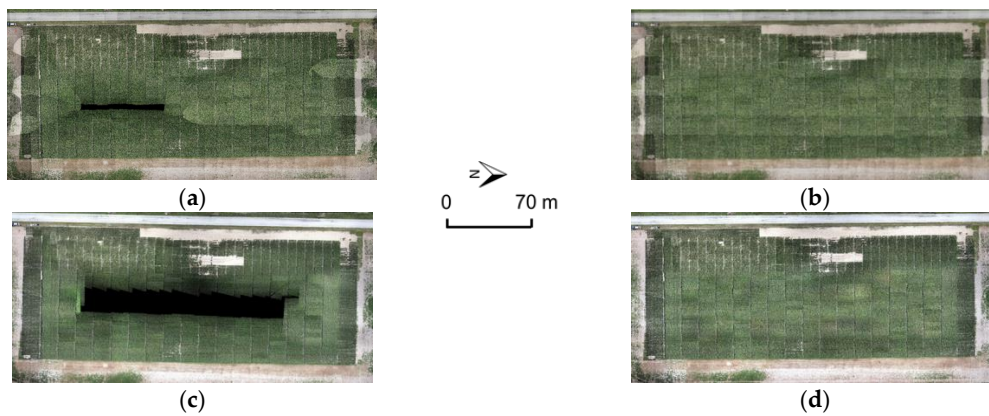


Figure 25. Atlanta-1 orthophotos generated through (a) traditional SfM, (b) fully GNSS/INS-assisted SfM, (c) Pix4D-1, and (d) Pix4D-2.

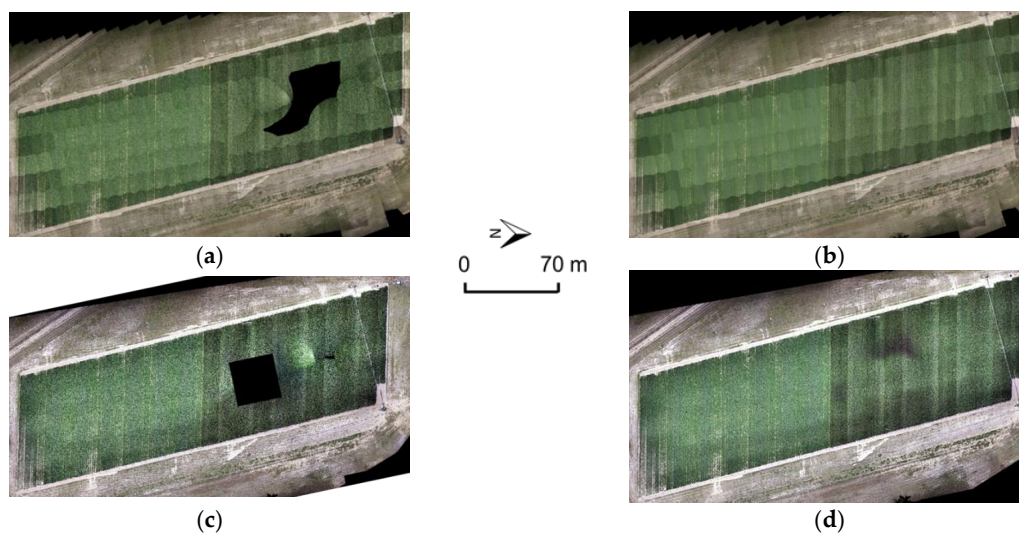


Figure 26. Atlanta-2 orthophotos generated through (a) traditional SfM, (b) fully GNSS/INS-assisted SfM, (c) Pix4D-1, and (d) Pix4D-2.

6. Conclusions and Recommendations for Future Work

In this paper, two frameworks, denoted as partially GNSS/INS-assisted SfM and fully GNSS/INS-assisted SfM, have been introduced for reliable aerial triangulation of UAV-based images captured over agricultural fields. The key motivation for such development is mitigating the limitations of existing SfM strategies, namely, poor distribution of derived object points and significant gaps in generated orthophotos when working with large image blocks over mechanized agricultural fields. The proposed strategies exploit the GNSS/INS trajectory in the image matching, ROP estimation, and bundle adjustment steps. At earlier stages of the SfM framework, the proposed approaches improve the matching reliability by using the GNSS/INS trajectory to limit the search space for conjugate points in overlapping images. For ROP estimation, a linear, iterative approach is used within the proposed strategies to refine trajectory-based ROPs while furnishing the ability to remove some matching outliers. In addition, the fully GNSS/INS-assisted framework eliminates the need for employing the time-consuming EOP recovery step by using the trajectory-based EOPs and handling remaining matching outliers through a RANSAC-based strategy. Finally, the two proposed frameworks augment the SfM strategy with a GNSS/INS-assisted bundle adjustment procedure that has a simpler implementation of the trajectory information in the adjustment while providing the possibility of refining the system calibration parameters. The performance of the proposed frameworks has been evaluated using eight datasets acquired over six agricultural fields. The comparison between the proposed strategies and a traditional SfM approach has shown the capability of the introduced frameworks to produce point clouds with better distribution and larger number of points while incorporating all the captured images in the SfM process without a significant increase in the processing time. Check point analysis shows centimeter level accuracy (i.e., RMSE value < 5 cm) of the reconstructed object points. Also, the derived point clouds for all the acquired datasets have shown a good alignment with LiDAR-based DSM at an overall precision range of ± 5 –10 cm. In terms of processing time, the fully GNSS/INS-assisted framework has been shown to be more efficient than the partially GNSS/INS-assisted one, even though the former have resulted in a larger number of object points. The performance of the proposed frameworks has also been compared with Pix4D Mapper Pro using two different settings: (1) “standard configuration” and (2) “accurate geolocation and orientation”, with the latter considering the GNSS/INS trajectory information after using the system calibration parameters in a preprocessing step to derive the image EOPs. The results demonstrate the superior performance of the proposed SfM frameworks in terms of number and distribution of the reconstructed object points and generating orthophotos without any gaps, when dealing with datasets with challenging texture conditions. In comparison with some commercial/opensource SfM alternatives that allow for the incorporation of prior position/orientation information, the introduced strategies provide the option of directly using GNSS/INS trajectory and refining system calibration parameters.

Since the repeatability and distinctiveness of the extracted local features play an important role in image-based 3D reconstruction, future work will focus on examining the performance of other local feature detection algorithms, such as SURF and accelerated KAZE (AKAZE) [74]. In addition, more robust matching outlier removal algorithms will be investigated. Conducting comparative analysis between the proposed SfM frameworks and other commercial/opensource triangulation software packages, such as PhotoScan and DroneDeploy, will be another focus of the future work.

Author Contributions: Conceptualization, S.M.H., T.Z. and A.H.; Data curation, S.M.H. and T.Z.; Formal analysis, S.M.H. and T.Z.; Methodology, S.M.H., T.Z. and A.H.; Supervision, A.H.; Writing—original draft, S.M.H. and T.Z.; Writing—review and editing, A.H. All authors have read and agreed to the published version of the manuscript.

Funding: 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), Army Research Office (ARO), 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.

Acknowledgments: Special acknowledgment is given to the Purdue TERRA team and the members of the Digital Photogrammetry Research Group (DPRG) for their work on system integration and data collections.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Gomiero, T.; Pimentel, D.; Paoletti, M.G. Is there a need for a more sustainable agriculture? *Crit. Rev. Plant Sci.* **2011**, *30*, 6–23. [[CrossRef](#)]
- Godfray, H.C.J.; Beddington, J.R.; Crute, I.R.; Haddad, L.; Lawrence, D.; Muir, J.F.; Toulmin, C. Food security: The challenge of feeding 9 billion people. *Science* **2010**, *327*, 812–818. [[CrossRef](#)]
- Sakschewski, B.; Von Bloh, W.; Huber, V.; Müller, C.; Bondeau, A. Feeding 10 billion people under climate change: How large is the production gap of current agricultural systems? *Ecol. Model.* **2014**, *288*, 103–111. [[CrossRef](#)]
- Wolfert, S.; Ge, L.; Verdouw, C.; Bogaardt, M.J. Big data in smart farming—A review. *Agric. Syst.* **2017**, *153*, 69–80. [[CrossRef](#)]
- Sedaghat, A.; Alizadeh Naeni, A. DEM orientation based on local feature correspondence with global DEMs. *GISci. Remote Sens.* **2018**, *55*, 110–129. [[CrossRef](#)]
- Aixia, D.; Zongjin, M.; Shusong, H.; Xiaoqing, W. Building Damage Extraction from Post-earthquake Airborne LiDAR Data. *Acta Geol. Sin. Engl. Ed.* **2016**, *90*, 1481–1489. [[CrossRef](#)]
- Mohammadi, M.E.; Watson, D.P.; Wood, R.L. Deep Learning-Based Damage Detection from Aerial SfM Point Clouds. *Drones* **2019**, *3*, 68. [[CrossRef](#)]
- Grenzdörffer, G.J.; Engel, A.; Teichert, B. The photogrammetric potential of low-cost UAVs in forestry and agriculture. *Int. Arch. Photogramm. Sens. Spat. Inf. Sci.* **2008**, *31*, 1207–1214.
- Ravi, R.; Hasheminasab, S.M.; Zhou, T.; Masjedi, A.; Quijano, K.; Flatt, J.E.; Habib, A. UAV-based multi-sensor multi-platform integration for high throughput phenotyping. In *Proceedings of the Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping IV*; International Society for Optics and Photonics: Baltimore, MD, USA, 2019; Volume 11008, p. 110080E.
- Shi, Y.; Thomasson, J.A.; Murray, S.C.; Pugh, N.A.; Rooney, W.L.; Shafian, S.; Rana, A. Unmanned aerial vehicles for high-throughput phenotyping and agronomic research. *PLoS ONE* **2016**, *11*, e0159781. [[CrossRef](#)] [[PubMed](#)]
- Yang, G.; Liu, J.; Zhao, C.; Li, Z.; Huang, Y.; Yu, H.; Zhang, R. Unmanned aerial vehicle remote sensing for field-based crop phenotyping: Current status and perspectives. *Front. Plant Sci.* **2017**, *8*, 1111. [[CrossRef](#)] [[PubMed](#)]
- Johansen, K.; Morton, M.J.; Malbeteau, Y.M.; Aragon, B.; Al-Mashharawi, S.K.; Ziliani, M.G.; Tester, M.A. Unmanned Aerial Vehicle-Based Phenotyping Using Morphometric and Spectral Analysis Can Quantify Responses of Wild Tomato Plants to Salinity Stress. *Front. Plant Sci.* **2019**, *10*, 370. [[CrossRef](#)] [[PubMed](#)]
- Santini, F.; Kefauver, S.C.; Resco de Dios, V.; Araus, J.L.; Voltas, J. Using unmanned aerial vehicle-based multispectral, RGB and thermal imagery for phenotyping of forest genetic trials: A case study in *Pinus halepensis*. *Ann. Appl. Biol.* **2019**, *174*, 262–276. [[CrossRef](#)]
- Lelong, C.; Burger, P.; Jubelin, G.; Roux, B.; Labbé, S.; Baret, F. Assessment of unmanned aerial vehicles imagery for quantitative monitoring of wheat crop in small plots. *Sensors* **2008**, *8*, 3557–3585. [[CrossRef](#)] [[PubMed](#)]
- Berni, J.A.; Zarco-Tejada, P.J.; Suárez, L.; Fereres, E. Thermal and narrowband multispectral remote sensing for vegetation monitoring from an unmanned aerial vehicle. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 722–738. [[CrossRef](#)]
- Hunt, E.R.; Hively, W.D.; Fujikawa, S.; Linden, D.; Daughtry, C.S.; McCarty, G. Acquisition of NIR-green-blue digital photographs from unmanned aircraft for crop monitoring. *Remote Sens.* **2010**, *2*, 290–305. [[CrossRef](#)]
- Zhao, J.; Zhang, X.; Gao, C.; Qiu, X.; Tian, Y.; Zhu, Y.; Cao, W. Rapid Mosaicking of Unmanned Aerial Vehicle (UAV) Images for Crop Growth Monitoring Using the SIFT Algorithm. *Remote Sens.* **2019**, *11*, 1226. [[CrossRef](#)]
- Masjedi, A.; Carpenter, N.R.; Crawford, M.M.; Tuinstra, M.R. Prediction of Sorghum Biomass Using Uav Time Series Data and Recurrent Neural Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Long Beach, CA, USA, 16–20 June 2019.

19. Zhang, X.; Zhao, J.; Yang, G.; Liu, J.; Cao, J.; Li, C.; Gai, J. Establishment of Plot-Yield Prediction Models in Soybean Breeding Programs Using UAV-based Hyperspectral Remote Sensing. *Remote Sens.* **2019**, *11*, 2752. [[CrossRef](#)]
20. Masjedi, A.; Zhao, J.; Thompson, A.M.; Yang, K.W.; Flatt, J.E.; Crawford, M.M.; Chapman, S. Sorghum Biomass Prediction Using Uav-Based Remote Sensing Data and Crop Model Simulation. In Proceedings of the IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 7719–7722.
21. Ravi, R.; Lin, Y.J.; Shamseldin, T.; Elbahnasawy, M.; Masjedi, A.; Crawford, M.; Habib, A. Wheel-Based Lidar Data for Plant Height and Canopy Cover Evaluation to Aid Biomass Prediction. In Proceedings of the IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 3242–3245.
22. Su, W.; Zhang, M.; Bian, D.; Liu, Z.; Huang, J.; Wang, W.; Guo, H. Phenotyping of Corn Plants Using Unmanned Aerial Vehicle (UAV) Images. *Remote Sens.* **2019**, *11*, 2021. [[CrossRef](#)]
23. Kitano, B.T.; Mendes, C.C.; Geus, A.R.; Oliveira, H.C.; Souza, J.R. Corn Plant Counting Using Deep Learning and UAV Images. *IEEE Geosci. Remote Sens. Lett.* **2019**, 1–5. [[CrossRef](#)]
24. Malambo, L.; Popescu, S.; Ku, N.W.; Rooney, W.; Zhou, T.; Moore, S. A Deep Learning Semantic Segmentation-Based Approach for Field-Level Sorghum Panicle Counting. *Remote Sens.* **2019**, *11*, 2939. [[CrossRef](#)]
25. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [[CrossRef](#)]
26. Bay, H.; Tuytelaars, T.; Van Gool, L. Surf: Speeded up robust features. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 404–417.
27. Fischler, M.A.; Bolles, R.C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* **1981**, *24*, 381–395. [[CrossRef](#)]
28. Ravi, R.; Lin, Y.J.; Elbahnasawy, M.; Shamseldin, T.; Habib, A. Simultaneous System Calibration of a Multi-LiDAR Multicamera Mobile Mapping Platform. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 1694–1714. [[CrossRef](#)]
29. Habib, A.; Zhou, T.; Masjedi, A.; Zhang, Z.; Flatt, J.E.; Crawford, M. Bore-sight Calibration of GNSS/INS-Assisted Push-Broom Hyperspectral Scanners on UAV Platforms. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 1734–1749. [[CrossRef](#)]
30. Khoramshahi, E.; Campos, M.B.; Tommaselli, A.M.G.; Vilijanen, N.; Mielonen, T.; Kaartinen, H.; Kukko, A. Accurate Calibration Scheme for a Multi-Camera Mobile Mapping System. *Remote Sens.* **2019**, *11*, 2778. [[CrossRef](#)]
31. LaForest, L.; Hasheminasab, S.M.; Zhou, T.; Flatt, J.E.; Habib, A. New Strategies for Time Delay Estimation during System Calibration for UAV-based GNSS/INS-Assisted Imaging Systems. *Remote Sens.* **2019**, *11*, 1811. [[CrossRef](#)]
32. Gabrlik, P.; Cour-Harbo, A.L.; Kalvodova, P.; Zalud, L.; Janata, P. Calibration and accuracy assessment in a direct georeferencing system for UAS photogrammetry. *Int. J. Remote Sens.* **2018**, *39*, 4931–4959. [[CrossRef](#)]
33. He, F.; Zhou, T.; Xiong, W.; Hasheminnasab, S.; Habib, A. Automated aerial triangulation for UAV-Based mapping. *Remote Sens.* **2018**, *10*, 1952. [[CrossRef](#)]
34. Fritz, A.; Kattenborn, T.; Koch, B. UAV-based photogrammetric point clouds-tree stem mapping in open stands in comparison to terrestrial laser scanner point clouds. In Proceedings of the ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Rostock, Germany, 4–6 September 2013; pp. 141–146.
35. Turner, D.; Lucieer, A.; Watson, C. An automated technique for generating georectified mosaics from ultra-high resolution unmanned aerial vehicle (UAV) imagery, based on structure from motion (SfM) point clouds. *Remote Sens.* **2012**, *4*, 1392–1410. [[CrossRef](#)]
36. Triggs, B.; McLauchlan, P.F.; Hartley, R.I.; Fitzgibbon, A.W. Bundle adjustment—A modern synthesis. In *International Workshop on Vision Algorithms*; Springer: Berlin/Heidelberg, Germany, 1999; pp. 298–372.
37. Rosten, E.; Drummond, T. Machine learning for high-speed corner detection. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 430–443.
38. Mikolajczyk, K.; Schmid, C. A performance evaluation of local descriptors. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1615–1630. [[CrossRef](#)]

39. Schmid, C.; Mohr, R.; Bauckhage, C. Evaluation of interest point detectors. *Int. J. Comput. Vis.* **2000**, *37*, 151–172. [[CrossRef](#)]
40. Karami, E.; Prasad, S.; Shehata, M. Image matching using SIFT, SURF, BRIEF and ORB: Performance comparison for distorted images. *arXiv* **2017**, arXiv:1710.02726.
41. Yi, K.M.; Trulls, E.; Lepetit, V.; Fua, P. Lift: Learned invariant feature transform. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2016; pp. 467–483.
42. Choy, C.B.; Gwak, J.; Savarese, S.; Chandraker, M. Universal correspondence network. In *Advances in Neural Information Processing Systems*; The MIT Press: Barcelona, Spain, 2016; pp. 2414–2422.
43. DeTone, D.; Malisiewicz, T.; Rabinovich, A. Superpoint: Self-supervised interest point detection and description. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 224–236.
44. Heymann, S.; Müller, K.; Smolic, A.; Froehlich, B.; Wiegand, T. SIFT implementation and optimization for general-purpose GPU. In Proceedings of the 15th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision, Plzen, Czech Republic, 29 January–1 February 2007.
45. Wu, C. SiftGPU: A GPU Implementation of Scale Invariant Feature Transform (SIFT) Method. Available online: <http://cs.unc.edu/~jccwu/siftgpu> (accessed on 1 July 2019).
46. Horn, B.K. Relative orientation. *Int. J. Comput. Vis.* **1990**, *4*, 59–78. [[CrossRef](#)]
47. Longuet-Higgins, H.C. A computer algorithm for reconstructing a scene from two projections. *Nature* **1981**, *293*, 133. [[CrossRef](#)]
48. Hartley, R.I. In defense of the eight-point algorithm. *IEEE Trans. Pattern Anal. Mach. Intell.* **1997**, *19*, 580–593. [[CrossRef](#)]
49. Zhang, Z. Determining the epipolar geometry and its uncertainty: A review. *Int. J. Comput. Vis.* **1998**, *27*, 161–195. [[CrossRef](#)]
50. Luong, Q.T.; Deriche, R.; Faugeras, O.; Papadopoulos, T. *On Determining the Fundamental Matrix: Analysis of Different Methods and Experimental Results*; Unit Ederechercheinria Sophiaantipolis: Valbonne, France, 1993.
51. Nistér, D. An efficient solution to the five-point relative pose problem. *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 756–777. [[CrossRef](#)]
52. Li, H.; Hartley, R. Five-point motion estimation made easy. In Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06), Hong Kong, China, 20–24 August 2006; Volume 1, pp. 630–633.
53. Cox, D.A.; Little, J.; O'shea, D. *Using Algebraic Geometry*; Springer Science Business Media: New York, NY, USA, 2006; Volume 185.
54. He, F.; Habib, A. Three-point-based solution for automated motion parameter estimation of a multi-camera indoor mapping system with planar motion constraint. *ISPRS J. Photogramm. Remote Sens.* **2018**, *142*, 278–291. [[CrossRef](#)]
55. Ortin, D.; Montiel, J.M.M. Indoor robot motion based on monocular images. *Robotica* **2001**, *19*, 331–342. [[CrossRef](#)]
56. Scaramuzza, D.; Fraundorfer, F.; Siegwart, R. Real-time monocular visual odometry for on-road vehicles with 1-point ransac. In Proceedings of the 2009 IEEE International Conference on Robotics and Automation, Kobe, Japan, 12–17 May 2009; pp. 4293–4299.
57. Hoang, V.D.; Hernández, D.C.; Jo, K.H. Combining edge and one-point ransac algorithm to estimate visual odometry. In *International Conference on Intelligent Computing*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 556–565.
58. He, F.; Habib, A. Automated relative orientation of UAV-based imagery in the presence of prior information for the flight trajectory. *Photogramm. Eng. Remote Sens.* **2016**, *82*, 879–891. [[CrossRef](#)]
59. Snavely, N.; Seitz, S.M.; Szeliski, R. Photo tourism: Exploring photo collections in 3D. In *ACM Transactions on Graphics (TOG)*; ACM: New York, NY, USA, 2006; Volume 25, pp. 835–846.
60. Dunn, E.; Frahm, J.M. Next Best View Planning for Active Model Improvement. In *BMVC*; The British Machine Vision Association: Oxford, UK, 2009; pp. 1–11.
61. Hartley, R.; Trunpf, J.; Dai, Y.; Li, H. Rotation averaging. *Int. J. Comput. Vis.* **2013**, *103*, 267–305. [[CrossRef](#)]
62. Martinec, D.; Pajdla, T. Robust rotation and translation estimation in multiview reconstruction. In Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17–22 June 2007; pp. 1–8.

63. Fitzgibbon, A.W.; Zisserman, A. Automatic camera recovery for closed or open image sequences. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 1998; pp. 311–326.
64. Haner, S.; Heyden, A. Covariance propagation and next best view planning for 3d reconstruction. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 545–556.
65. Cornelis, K.; Verbiest, F.; Van Gool, L. Drift detection and removal for sequential structure from motion algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 1249–1259. [[CrossRef](#)] [[PubMed](#)]
66. Govindu, V.M. Combining two-view constraints for motion estimation. In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2001, Kauai, HI, USA, 8–14 December 2011; Volume 2, p. II.
67. Chatterjee, A.; Madhav Govindu, V. Efficient and robust large-scale rotation averaging. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, NSW, Australia, 1–8 December 2013; pp. 521–528.
68. Sinha, S.N.; Steedly, D.; Szeliski, R. A multi-stage linear approach to structure from motion. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 267–281.
69. Arie-Nachimson, M.; Kovalsky, S.Z.; Kemelmacher-Shlizerman, I.; Singer, A.; Basri, R. Global motion estimation from point matches. In Proceedings of the 2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission, Zurich, Switzerland, 13–15 October 2012; pp. 81–88.
70. Cui, Z.; Jiang, N.; Tang, C.; Tan, P. Linear global translation estimation with feature tracks. *arXiv* **2015**, arXiv:1503.01832.
71. He, F.; Habib, A. Target-based and Feature-based Calibration of Low-cost Digital Cameras with Large Field-of-view. In Proceedings of the ASPRS 2015 Annual Conference, Tampa, FL, USA, 4–8 May 2015.
72. Habib, A.; Xiong, W.; He, F.; Yang, H.L.; Crawford, M. Improving orthorectification of UAV-based push-broom scanner imagery using derived orthophotos from frame cameras. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *10*, 262–276. [[CrossRef](#)]
73. Lin, Y.C.; Cheng, Y.T.; Zhou, T.; Ravi, R.; Hasheminasab, S.M.; Flatt, J.E.; Habib, A. Evaluation of UAV LiDAR for Mapping Coastal Environments. *Remote Sens.* **2019**, *11*, 2893. [[CrossRef](#)]
74. Alcantarilla, P.F.; Solutions, T. Fast explicit diffusion for accelerated features in nonlinear scale spaces. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *34*, 1281–1298.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).