Automated Aerial Triangulation for UAV-Based Mapping

Fangning He *, **Tian Zhou, Weifeng Xiong, Seyyed Meghdad Hasheminnasab and Ayman Habib**

Lyles School of Civil Engineering, Purdue University, West Lafayette, IN 47907, USA;
zhou732@purdue.edu (T.Z.); weifengxiong1992@gmail.com (W.X.); hashemin@purdue.edu (S.M.H.);
ahabib@purdue.edu (A.H.)

*** Correspondence: fangninhe@gmail.com Tel.: +1-765-479-5945

check for
updates

Abstract: Accurate 3D reconstruction/modelling from unmanned aerial vehicle (UAV)-based imagery has become the key prerequisite in various applications. Although current commercial software has automated the process of image-based reconstruction, a transparent system, which can be incorporated with different user-defined constraints, is still preferred by the photogrammetric research community. In this regard, this paper presents a transparent framework for the automated aerial triangulation of UAV images. The proposed framework is conducted in three steps. In the first step, two approaches, which take advantage of prior information regarding the flight trajectory, are implemented for reliable relative orientation recovery. Then, initial recovery of image exterior orientation parameters (EOPs) is achieved through either an incremental or global approach. Finally, a global bundle adjustment involving Ground Control Points (GCPs) and check points is carried out to refine all estimated parameters in the defined mapping coordinate system. Four real image datasets, which are acquired by two different UAV platforms, have been utilized to evaluate the feasibility of the proposed framework. In addition, a comparative analysis between the proposed framework and the existing commercial software is performed. The derived experimental results demonstrate the superior performance of the proposed framework in providing an accurate 3D model, especially when dealing with acquired UAV images containing repetitive pattern and significant image distortions.

Keywords: unmanned aerial vehicle; 3D reconstruction; structure from motion; relative orientation; exterior orientation parameters; bundle adjustment

1. Introduction

In the past few years, low-cost Unmanned Airborne Vehicles (UAVs) equipped with consumer-grade imaging systems (e.g., commercial off-the-shelf digital camera) have emerged as a potential remote sensing platform that could satisfy the needs of a wide range of applications, such as precision agriculture [1–9], environmental monitoring [10–14], forest inventory [15,16], wildlife research [17,18], and archaeological applications [19,20]. Compared to conventional human-operated terrestrial and airborne mapping/remote sensing systems, the advantages of UAVs include their low-cost, small size, low flying height, ease of storage and deployment, and the capability of providing high spatial resolution geospatial data at a higher data collection rate [21–23]. From a mapping point of view, deriving accurate three-dimensional (3-D) geo-spatial information from UAV-based imagery requires the interior orientation parameters (IOPs) of the utilized camera, and exterior orientation parameters (EOPs) of the involved images. The IOPs, which encompass the internal sensor characteristics such as focal length and camera-specific distortions, can be derived

from a camera calibration process [24,25]. The EOPs, which defines the position and orientation of the imaging system at the moments of exposure, can be either derived through an indirect or a direct geo-referencing process. For indirect geo-referencing, the image EOPs are traditionally established using ground control points (GCPs) within a bundle adjustment (BA) procedure. However, the set-up of control points are time-consuming and costly activities. Alternatively, thanks to the available GNSS/INS Position and Orientation Systems (POS) onboard, direct geo-referencing simplifies the derivation of the image EOPs without the need for any GCPs [26–28]. Unfortunately, due to the limited endurance and payload constraints, current low-cost UAV-based mapping systems are usually equipped with consumer-grade geo-referencing units and light-weight imaging systems. Compared to survey-grade GNSS/INS units, the consumer-grade systems are relatively small and provide inaccurate position and orientation information. At present, Structure from Motion (SfM), which was initiated by the computer vision research community, has been widely used for automated triangulation of overlapping UAV-based frame imagery while using minimal GCPs and/or low-quality navigation information from consumer-grade GNSS/INS units [29]. Similar to the aerial triangulation procedure that has been adopted by the photogrammetric community for decades, SfM is usually implemented in three steps to simultaneously estimate the EOPs of the involved images and derive 3D coordinates of matched features within the overlapping area [30]. In the first step, the relative orientation parameters (ROPs) relating stereo-images are initially estimated using automatically identified conjugate point and/or line features [31]. Then, a local reference coordinate system is established to define an arbitrary datum for deriving the image EOPs as well as 3D coordinates of matched points. Finally, a bundle adjustment procedure is implemented to refine the EOPs and object coordinates derived in the second step. Current commercial software (e.g., Pix4D, PhotoScan, etc.) has automated the SfM process for UAV image-based 3D reconstruction. However, for some emerging applications, such as precision agriculture, accurate UAV-based mapping remains a challenging task. This is mainly due to the fact that the acquired imagery usually contains poor and/or repetitive texture, which can severely impact the relative orientation recovery for the involved stereo-pairs. In addition, due to the black-box nature of commercial software, it is always hard to figure out reasons for the internal failure of processing. In this regard, this paper aims at proposing a transparent processing framework which can be augmented with different user-defined constraints, for the automated aerial triangulation of UAV-based imagery. To be more specific, this research will be focusing on the following issues:

- Automated relative orientation recovery of UAV-based images in the presence of prior information regarding the flight trajectory,
- Initial recovery of image EOPs through either incremental or global SfM-based strategies,
- Accuracy analysis of the derived 3D reconstruction through check point analysis, and
- Comparison of the proposed approach against available commercial software, such as Pix4D.

To address these issues, the theoretical background for the proposed framework is introduced in the next section. Then, the utilized methodology is explained. Afterwards, experimental results with real datasets are discussed. Finally, drawn conclusions as well as recommendations for future work are presented.

2. Theoretical Background

This section introduces the theoretical background for the automated aerial triangulation and SfM-based 3D reconstruction. First, the mathematical model and closed-form solutions for relative orientation is introduced. Then, a literature review of existing research efforts regarding the recovery of image EOPs is given. Finally, the concept of bundle adjustment, which is usually conducted as the final refinement for image-based 3D reconstruction, is presented.

2.1. Relative Orientation Recovery

Accurate estimation of ROPs is a prerequisite for image-based 3D reconstruction, which follows an SfM-based framework. For a given stereo-pair, ROP estimation involves the derivation of five parameters, which include three rotation angles and two translation parameters (i.e., an arbitrary scale is assumed for the ROP estimation procedure). The most well-known approach for ROP recovery is based on the co-planarity constraint [32], where a least-squares solution is derived while using a minimum of five conjugate points. As shown in Figure 1, the co-planarity constraint describes the fact that an object point P, conjugate image points p_1 and p_2, and the two perspective centers O_1 and O_2 of a stereo-pair must lie on the same plane. In the mathematical expression for the co-planarity constraint as presented in Equation (1), p_1 and p_2 are the two conjugate image points, where $p = (x, y, -c)$ represents the image coordinates corrected for the principal point offset and camera-specific distortions. The rotation matrix R, which is defined by the three rotation angles ω, ϕ, and κ, describes the relative rotation relating the two stereo-images. $\vec{T}$ is the translation vector describing the baseline between the stereo-images, and it can be defined by three translation parameters (T_x, T_y, T_z). The symbol $\times$ denotes the cross product between two vectors. Due to the nonlinear nature of the co-planarity model, the least-squares solution requires approximate initial values for the unknown parameters. Then, these parameters are refined through an iterative process until a pre-defined stopping criterion is satisfied (e.g., insignificant changes can be observed between successive estimates of the parameters). However, establishing good approximations are not always possible, especially when the mapping platform exhibits excessive maneuvers between the data acquisition epochs (e.g., close range mapping applications).

$$p_1 \cdot \left(\vec{T} \times R p_2 \right) = 0 \tag{1}$$
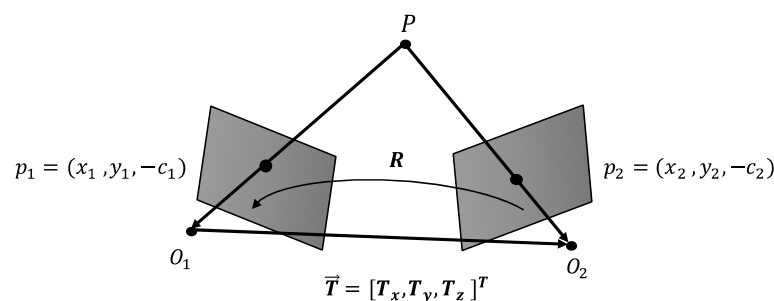

Figure 1. The co-planarity model relating stereo-images.

To date, several closed-form solutions (e.g., the eight-point and the five-point algorithms), which do not require approximations, have been developed for ROP recovery [33–35]. These solutions are based on the concept of the Essential matrix, which is derived from the co-planarity constraint and encapsulates the epipolar geometry relating stereo-images. Since the cross product of two vectors can be expressed as a matrix-vector multiplication, Equation (1) can be simplified as Equation (2) using the 3-by-3 skew-symmetric matrix $\hat{T}$. Then, according to Equation (2), one can derive the expression of the Essential matrix as shown in Equation (3). It is worth noting that the nine elements of the Essential matrix are defined by the five elements of the ROPs (i.e., three rotation angles and two translation components). Therefore, there must be four additional constraints that can be imposed on the nine elements of the Essential matrix E [31]. Given that the Essential matrix has rank two, the first cubic constraint on the nine unknown parameters of the Essential matrix is presented as in Equation (4), where the determinant of the matrix has to be zero. Then, another two constraints—namely the trace constraints—are deduced from the equality

as established in Equation (5). Finally, the fourth constraint is that the nine elements of the Essential matrix can be only determined up to a scale.

$$p_1^T \hat{T} R p_2 = 0 \text{ where, } \hat{T} = \begin{bmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix} \tag{2}$$

$$E = \hat{T}R = \begin{bmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix} R = \begin{bmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & e_{33} \end{bmatrix} \tag{3}$$

$$\det(E) = 0 \tag{4}$$

$$EE^TE - \frac{1}{2}trace\left(EE^T\right)E = 0 \tag{5}$$

The first closed-form solution of the Essential matrix is proposed by Longuet–Higgins for recovering the structure of a scene from two views that have been captured by a calibrated camera [33]. In spite of its simplicity, this approach fails to consider both the cubic and trace constraints as shown in Equations (4) and (5), respectively. Therefore, a minimum of eight conjugate points is required in this approach, and it has been criticized for its excessive sensitivity to noise in the image coordinates of conjugate point pairs as well as having an object space that is almost planar. An improvement to such an eight-point algorithm was proposed by Hartley [34], where a coordinate normalization procedure is applied to bring the origin of the image coordinate system to the centroid of the involved points. Experimental results from Hartley's work demonstrated that with image coordinate normalization, the performance of the eight-point algorithm is almost at the same quality as the iterative nonlinear approach. Given that a minimum of five conjugate point pairs are sufficient for ROP recovery, several five-point algorithms [35–41] have been proposed as alternatives to the eight-point algorithm. To date, the most well-known five-point algorithm is the one proposed by Nistér [35], which is based on a modified Gaussian–Jordan elimination procedure. An improvement for Nistér's five-point algorithm is proposed by Li and Hartley while using a hidden variable resultant approach for the estimation of unknown parameters [41]. This improvement is easier to understand and implement. However, it can be much more computationally expensive when compared to the original Nistér's approach. Since the five-point algorithms enforce all inherent constraints among the elements of the Essential matrix, they are capable of providing more accurate estimate of ROPs when compared to eight-point approaches, especially when dealing with noisy conjugate image measurements or acquired images from planar scenes [37].

The above-mentioned ROP recovery procedures are based on reliable conjugate points in stereo imagery. However, illumination changes, induced occlusions by perspective geometry, and arising ambiguity from repetitive patterns will introduce outliers in automatically identified conjugate features. For robust ROP estimation, it is always necessary to augment these closed-form solutions (i.e., eight or five-point algorithms) with strategies for outlier removal. One of the commonly-used strategies to filter outliers for ROP recovery is Random Sample Consensus (RANSAC) [42]. In practice, RANSAC starts with conducting random draws of the necessary samples to derive an initial estimate of the Essential matrix. Then, the point-to-epipolar line distances are evaluated for all available matches according to the co-planarity model and the estimated Essential matrix. Finally, the draw that results in the largest consensus is used together with the compatible matches to derive a reliable estimate of the ROPs. Despite its potential, RANSAC would require an excessive number of trials when dealing with scenarios that require large samples and/or have high percentage of outliers. Moreover, RANSAC might fail to provide a set of matches that supports correct ROP estimate when there is a false hypothesis providing larger consensus.

In order to mitigate the impact of RANSAC drawbacks, some approaches, which take advantage of the availability of additional constraints on the system trajectory during data acquisition, have been developed to reduce the number of required feature correspondences for ROP recovery [43–46]. These approaches, which were mainly initiated by the mobile robotics research community, assume one or more parameters of the stereo-based relative orientation to be known. For example, considering the fact that the relative rotation between stereo-images can be alternatively defined by a rotation angle around an axis (i.e., reference direction) in space [47], several three-plus-one algorithms [48–50], which utilize three point correspondences and a known rotation reference direction, have been developed as a substitute of the classic five-point algorithm. In practice, prior information regarding the reference direction can be either derived from a detected vanishing point [51] or using a gravity sensor onboard mobile mapping platforms, where the gravity vector becomes the reference direction [44]. A recent three-point solution has been proposed by Fraundorfer et al. [44]. In his work, a simplified Essential matrix is estimated from three-point correspondences using two known rotation angles, which are acquired from a Smart-phone. In addition to these three-point solutions, another example of using prior trajectory information to facilitate the ROP recovery process is introduced by Troiani [46]. In Troiani's work, a two-point algorithm has been proposed for estimating the translation components of the ROPs while relying on available rotation angles relating consecutive images from an Inertial Measurement Unit (IMU), which has been rigidly attached to a monocular camera. It is worth noting that the three translation parameters (T_x, T_y, T_z) are linearly recovered to an arbitrary scale through two-point correspondences in this approach. In terms of UAV-based mapping, He and Habib [23,52] proposed a two-point approach for automated relative orientation recovery while considering prior information regarding the flight trajectory. This approach assumes that the UAV platform is moving at a constant flying height while operating a nadir-looking camera. The derived experimental results from different real datasets have demonstrated the feasibility of this two-point approach in providing reliable ROPs from UAV-based imagery in the presence of a high percentage of matching outliers. In this research, the two-point solution is adopted for the proposed UAV-based aerial triangulation procedure.

2.2. Exterior Orientation Estimation

Now that the ROPs among the overlapping imagery are estimated, the SfM-based framework generally adopts either an incremental or global strategy to establish the EOPs of the involved images. These estimated EOPs can be finally utilized as input values in the bundle adjustment for additional refinement. Existing incremental approaches usually estimate the EOPs through an image augmentation process. For example, Snavely [53] proposed an incremental SfM procedure for 3D reconstruction using Internet images. In this procedure, a reference frame is initially established from a single pair of images that has a large number of matched points and a long baseline. Then, the remaining images are sequentially added to the reference frame based on the number of feature correspondences with previously referenced images. The Direct Linear Transform (DLT) [30] is incorporated within a RANSAC procedure for deriving the EOPs of the augmented images [53]. Another incremental approach to recover the EOPs for either a closed-loop or open-sequence of acquired images is developed by Fitzgibbon and Zisserman [54]. In their approach, the relative orientation parameters are first recovered through trifocal tensors [55] for all consecutive image triplets. Afterwards, an incremental approach is applied to gradually integrate the image triplets to subsets. Finally, these subsets are augmented into a single image block. Although incremental SfM has been widely utilized for various 3D reconstruction applications while using ordered/unordered images [56,57], the high time complexity, which is commonly known to be $O(n^4)$ for a collection of n images, impedes the widespread adoption of such a simple strategy for large image datasets [58]. In addition, it is worth noting that the selection of initial image pair/triplet can be critical for incremental SfM [59]. In practice, due to the increased redundancy, initialization from a location with an adequate number of overlapping images usually leads to more robust parameter estimation.

On the other hand, an initial stereo-pair with insufficient image matches may result in unreliable 3D reconstruction as well as the failure of image augmentation. The performance of the incremental algorithms also depends on the order of augmented images. According to the existing body of literature [60], a designed image augmentation which considers the geometric compatibility among overlapping images can be adopted to mitigate the impact of error propagation for reliable estimation of image EOPs. Moreover, intermediate bundle adjustment, which is periodically conducted during the image augmentation process, is another commonly-used technique in most incremental SfM approaches. Although such intermediate bundle adjustment ensures successful augmentation of the individual images into the final image block, it can be computationally expensive.

Different from the incremental algorithms, a global SfM strategy aims at simultaneously establishing the EOP estimation for all involved images while providing better efficiency and accuracy. Currently, most of the state-of-the-art global approaches are based on a two-step strategy. Specifically, in the first step, a multiple rotation averaging procedure is utilized to simultaneously solve image orientations using all the derived ROPs. Then, the positional components are derived through a global translation averaging while using the estimated image rotations. According to Hartley [61], given a set of m images with n available stereo-based relative rotations (e.g., R_i^j), the multiple rotation averaging aims at finding the m optimum global rotation estimates (e.g., R_i^{global} and R_j^{global}) for all the involved images while satisfying the n compatibility constraints in the form of $R_i^j = (R_j^{global})^T R_i^{global}$. To date, several approaches have been developed for solving the multiple rotation averaging problem. For example, Martinec and Pajdla [62] have demonstrated that it is possible to derive a closed-form solution for all the rotation matrices through a singular value decomposition (SVD) on the set of linear equations for the established compatibility constraints. However, it is worth noting that such SVD-based solution fails to consider the inherent orthogonality constraints among the nine elements of the rotation matrix. On the other hand, considering the fact that a rotation in 3D space can be represented in different forms (e.g., Euler angles or quaternions), Martinec and Pajdla introduced another solution for the multiple rotation averaging while using quaternions. Instead of having nine elements in a rotation matrix, a quaternion gives a concise way to represent a rotation in 3D space through four numbers that represent a unit vector. Therefore, the utilization of quaternions can significantly reduce the number of utilized equations for the estimation of rotations. Unfortunately, at this time, there is no satisfactory way to derive a linear solution from the quaternion-based approach while enforcing the unit length constraint on the resulting quaternion [62]. Instead of using Euler angles or quaternions, recent research efforts, which are based on Lie-algebra representations and robust L_1 optimization, have demonstrated better performance for the multiple rotation averaging [63]. Interested readers can refer to Hartley [61] and Carlone et al. [64] for more information about the theory of multiple rotation averaging.

In contrast to multiple rotation averaging, the estimation of translational components for all the available images can be more challenging since the derived stereo-based translations are only determined up to an arbitrary scale. The existing research efforts for the global translation estimation include some linear approaches [65–68], which are mainly based on the compatibility constraint among different translation vectors. For example, Govindu [65] estimated the positions of a group of images relative to a common reference coordinate system by enforcing consistency constraints among stereo-based translation directions. However, this approach cannot deal with images captured in a linear-trajectory configuration. Different from Govindu's approach, Sinha et al. [67] determined the global translations through stereo-based registration. However, such a registration process requires tracking of conjugate 3D points reconstructed in all possible stereo-pairs. Arie-Nachimson et al. [68] introduced another solution for global translation estimation while using a novel decomposition of the Essential matrix. This approach can deal with stereo-pairs with different baseline lengths. However, it still suffers from the degeneracy caused by linear trajectory configuration. In order to resolve such degeneracy in translation estimation, Cui et al. [69,70]

utilized corresponding image points, which are derived through a feature tracking process, to establish an absolute scale for all the translation parameters. However, careful outlier detection/removal is required in the feature tracking process. On the other hand, considering the fact that a common scale can be determined within an image-triplet, some trifocal tensor-based approaches have been investigated for global translation estimation. Recently, Jiang et al. [71] proposed a novel linear constraint for image triplets, and derived position estimates for all the available images through a least-squares adjustment. Such trifocal tensor-based approach is capable of dealing with the degenerate camera motion problem (i.e., linear trajectory). However, strong connection among overlapping images (i.e., an adequate number of favorably distributed point correspondences) is usually required. Another draw-back of the trifocal tensor-based approach is that for a dataset with huge number of images, the total number of available image triplets can be significantly greater than the number of stereo-pairs, which consequently leads to a low computational efficiency. Although the global strategy has recently attracted more attention in both photogrammetric and computer vision research communities, the incremental SfM is still the most commonly-used approach in the existing commercial software. Therefore, this paper is dedicated to investigate the performance of both incremental and global strategies for UAV image-based 3D reconstruction by presenting a transparent framework for automated aerial triangulation.

2.3. Bundle Adjustment

In existing photogrammetric triangulation/SfM approaches, bundle adjustment (BA) is a commonly-used process to simultaneously refine the 3D coordinates of the scene points, the EOPs of the involved images, and/or the IOPs of the utilized cameras [72]. The classic photogrammetric bundle adjustment is based on the well-known collinearity equations. It can be formulated as a nonlinear least-squares problem, which aims at minimizing the total back-projection error between the observed image point coordinates and predicated feature locations [73]. In recent years, bundle adjustment has been further expanded to deal with a wide variety of situations, such as the utilization of different features (e.g., line [25,74], curves [75], etc.), the reconstruction of dynamic scene objects [76], and the employment of non-quadratic error models [77]. Interested readers can refer to the review conducted by Triggs et al. [77] for more details regarding modern bundle adjustment techniques [78,79].

3. Methodology

In this section, the proposed framework for the automated aerial triangulation is introduced. Similar to most existing procedures for UAV image-based 3D reconstruction, the proposed framework is accomplished through three steps. In the first step, the ROPs relating stereo-images are directly derived from conjugate point features. In order to deal with UAV-based imagery acquired in the presence of a high percentage of matching outliers, two approaches, which exploit prior information regarding the flight trajectory, are adopted. In the second step, the initial recovery of image EOPs is achieved. More specifically, the proposed framework investigates both incremental and global strategies for the EOPs recovery of all the involved imagery. Finally, in the third step, a global bundle adjustment, which is able to integrate both GCPs and Check Points, is carried out for indirect geo-referencing and accuracy analysis. Figure 2 illustrates the workflow of the proposed procedure.

3.1. Automated Relative Orientation

Considering the fact that current UAV-based data acquisition is usually executed according to a mission plan while relying on a consumer-grade navigation sensor within the platform's autopilot, the two approaches developed by He and Habib [23]—namely two-point and iterative five-point approaches—have been adopted in the proposed framework for automated relative orientation. In this research, SIFT (scale-invariant feature transform) detector and descriptors [80] are utilized to derive initial matches among overlapping images.
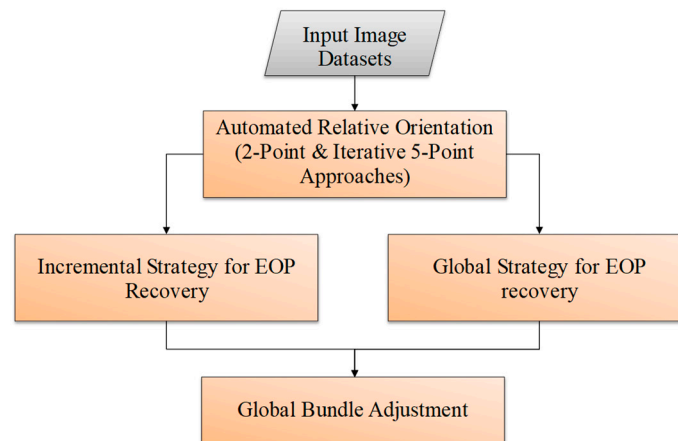
Figure 2. The proposed framework for the automated aerial triangulation of unmanned aerial vehicle (UAV)-based imagery.

The two-point approach is based on a common flight configuration for UAV-based mapping, in which the platform is moving at a constant flying height while operating a nadir-looking camera (i.e., we are dealing with vertical images that have been captured from the same flying height). Such flight configuration is commonly known as "planar motion", where the platform is constrained to move on a horizontal plane, and the rotation of the camera is constrained along an axis orthogonal to the horizontal plane. As shown in Figure 3, such UAV-based planar motion leads to two geometric constraints to simplify the estimation of relative orientation parameters among the image stereo-pairs. The two geometric constraints include:

- The rotation angle ω and ϕ between overlapping stereo-images can be assumed to be zero, and
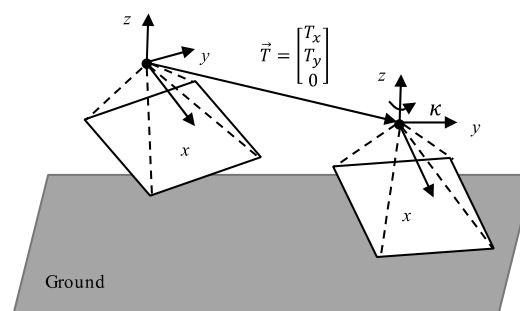- The translation component T_z is approximated to be zero.

Figure 3. Stereo-images from a UAV platform equipped with a nadir-looking camera while moving at a constant flying height.

Based on the two geometric constraints, the rotation matrix R and the translation vector $\vec{T}$ relating the stereo-images can be expressed as in Equation (6). Then, substituting both R and $\vec{T}$ into Equation (3) leads to the simplified Essential matrix in Equation (7), where $L_1, L_2, L_3,$ and L_4 are the four unknown elements within the simplified Essential matrix. It is worth noting that $L_1, L_2, L_3,$ and L_4 are derived from three independent parameters (T_x, T_y, and κ). Therefore, there should be one more constraint relating the four elements of the simplified Essential matrix. Through a closer inspection of Equation (7), one can derive the additional constraint as presented in Equation (8). In addition, since $L_1, L_2, L_3,$ and L_4 can be only determined up to an arbitrary scale, two pairs of conjugate points are sufficient for deriving the simplified Essential matrix through a closed form solution. Similar to the conventional

five/eight-point algorithms, the two-point approach can be incorporated within a RANSAC framework for outlier removal.

$$R = \begin{bmatrix} \cos\kappa & -\sin\kappa & 0 \\ \sin\kappa & \cos\kappa & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } \vec{T} = \begin{bmatrix} T_x \\ T_y \\ 0 \end{bmatrix} \tag{6}$$

$$
\begin{aligned}
E = \hat{T}R &= \begin{bmatrix} 0 & 0 & Ty \\ 0 & 0 & -Tx \\ -Ty & Tx & 0 \end{bmatrix} \begin{bmatrix} \cos\kappa & -\sin\kappa & 0 \\ \sin\kappa & \cos\kappa & 0 \\ 0 & 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} 0 & 0 & T_y \\ 0 & 0 & -T_x \\ -T_y\cos\kappa + T_x\sin\kappa & T_y\sin\kappa + T_x\cos\kappa & 0 \end{bmatrix} \\
&= \begin{bmatrix} 0 & 0 & L_1 \\ 0 & 0 & L_2 \\ L_3 & L_4 & 0 \end{bmatrix}
\end{aligned}
\tag{7}
$$

$$L_1^2 + L_2^2 - L_3^2 - L_4^2 = 0 \tag{8}$$

The iterative five-point approach starts from the co-planarity model while assuming the availability of prior information regarding the platform trajectory between the images of a stereo pair. Given approximate values for the platform's rotation matrix R and translation $\vec{T}$ between the images of a stereo-pair and assuming unknown incremental rotation and translation corrections (δR and δT), the co-planarity model can be modified to the form in Equation (9). As shown in Equation (9), $\hat{T}$ is the 3-by-3 skew-symmetric matrix determined by the approximate values for the translation parameters T_x, T_y, and T_z, and δT is the 3-by-3 skew-symmetric matrix comprised from the unknown corrections to the approximate translation vector. R is the rotation matrix defined by the approximate angles ω, ϕ, and κ, which can be derived from either the assumed flight trajectory or the measurements from onboard GNSS/INS unit. δR describes the unknown incremental rotation matrix defined by the incremental angles $\Delta\omega$, $\Delta\phi$, and $\Delta\kappa$. In practice, one translation correction (e.g., ΔT_x) can be set to 0 since the translation is only determined up to an arbitrary scale. Moreover, assuming that the deviations from the approximate rotation R are small (i.e., $\Delta\omega$, $\Delta\phi$, and $\Delta\kappa$ are small rotation angles), the incremental rotation matrix δR can be represented as in Equation (10). Substituting Equation (10) into Equation (9) and ignoring the second-order correction terms lead to a linear equation in five unknown parameters (i.e., two translation corrections and three incremental angles). Given five or more conjugate point pairs, one can derive a least-squares solution for the unknown corrections. In the iterative five-point approach, the derived corrections are used to refine the approximate ROPs through an iterative procedure until a convergence criterion is achieved. A built-in outlier removal process is adopted within the iterative procedure by imposing constraints on the normalized image coordinates according to epipolar geometry. Different from the two-point approach that assumes a planar motion of the utilized imaging system, the iterative five-point approach is capable of dealing with acquired UAV imagery at any tilt angles and vertical translation. However, accurate initial approximations, which are close to the true values of ROPs in stereo-pairs, are always required.

$$p_1^T\left(\hat{T} + \delta\hat{T}\right)\delta R R p_2 = 0 \tag{9}$$

$$\delta R = \begin{bmatrix} 1 & -\Delta\kappa & \Delta\phi \\ \Delta\kappa & 1 & -\Delta\omega \\ -\Delta\phi & \Delta\omega & 1 \end{bmatrix} \tag{10}$$

In practice, to deal with the acquired UAV images that exhibit significant variations from the designed flight plan (e.g., when operating a light-weight UAV in a relatively windy condition),

a hybrid strategy which integrates both the two-point and iterative five-point approaches can be adopted. More specifically, 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. Interested readers can refer to He and Habib [23] for more information about the two-point and iterative five-point approaches.

3.2. Incremental Strategy for EOP Recovery

Now that the ROPs among overlapping imagery are estimated, either an incremental or global strategy can be incorporated into the proposed automated aerial triangulation framework to establish the EOPs of the involved images. In this section, the incremental approach for the initial recovery of image EOPs is presented. The proposed approach is initiated by using seed images to define a local reference coordinate system. Then, the remaining images are sequentially augmented into the final image block or trajectory through closed-form solutions for both the rotational and positional components of the EOPs.

3.2.1. Local Reference Coordinate System Initialization

The proposed incremental approach starts with selecting an initial image triplet, which is comprised of three overlapping images, to define the local reference coordinate system. In order to find the optimum candidate for the initial image triplet, two conditions have to be satisfied:

- There should be a sufficient number of feature correspondences within the selected image triplet, and
- There should be a good geometric configuration among the three overlapping images.

In this research, the first condition is easily satisfied by maximizing the total number of conjugate points within the image triplet. The second condition, on the other hand, can be achieved through a compatibility analysis, which aims at evaluating the geometric configuration within the selected image triplet. Before presenting the proposed compatibility analysis, we first introduce the geometric constraints within an image triplet. Figure 4 depicts a sample image triplet comprised of three images i, j, and k. As can be seen in the figure, $\left(r_j^i,\ R_j^i\right)$, $\left(r_k^i,\ R_k^i\right)$, and $\left(r_k^j,\ R_k^j\right)$ are the relative orientation of the three involved stereo-pairs $(i,\ j)$, $(i,\ k)$, and $(j,\ k)$ within the image triplet, respectively. Given that the stereo-based translation components (i.e., r_j^i, r_k^i, and r_k^j) are only recovered to an arbitrary scale, a mathematical model to define the common scale within the image triplet has been provided in Equation (11). The conceptual basis of this mathematical model is based on the fact that within an image triplet, one relative translation vector can be expressed as a summation of the other two while considering appropriate scale factors. As shown in Equation (11), the common scale within the image triplet (i, j, k) is defined by r_j^i, which is the translation vector relating the two images i and j. λ_1 and λ_2 are two unknown scale factors for the other two translations r_k^i and $R_j^i r_k^j$. For each image triplet, a system of three linear equations in the two unknowns λ_1 and λ_2 can be established. One can derive a solution for the two unknown scale factors through a classic least-squares approach.

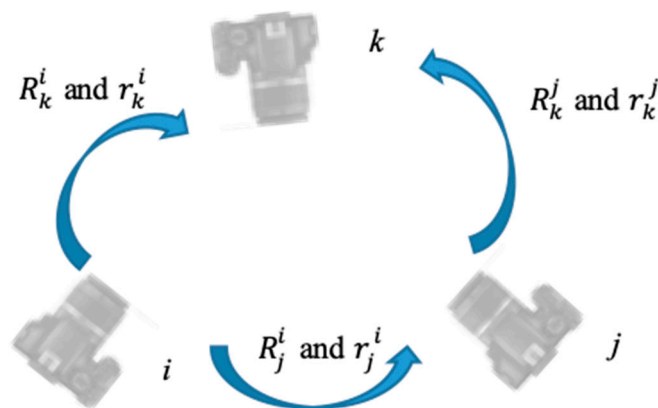$$r_j^i\ =\ \lambda_1 r_k^i - \lambda_2 R_j^i r_k^j \tag{11}$$

Figure 4. Image triplet for establishing the local coordinate system.

Once the values for the two scale factors λ_1 and λ_2 are determined, the EOPs of each image within the image triplet can be recovered relative to the local reference frame, which is established as the camera coordinate system of image i. Such a local reference frame will be denoted as l in this research. With a closer inspection of the image triplet shown in Figure 4, one can come up with two sets of EOP estimates of image k through the EOPs of either image i or j. The mathematical expressions for deriving the two different EOP estimates are presented in Equations (12) and (13), respectively. As can be seen in Equation (12), the first EOP estimate of image k, which is represented as $r^l_{k[i]}$ and $R^l_{k[i]}$, only relies on the scale factor λ_1 and the ROPs relating images i and k. Alternatively, the second estimate—$r^l_{k[j]}$ and $R^l_{k[j]}$—is derived through the EOPs of image j while using the scale factor λ_2 and the ROPs within stereo-pair (j, k).

$$R^l_{k[i]} \;=\; R^i_k \quad \text{and} \quad r^l_{k[i]} \;=\; \lambda_1 r^i_k \tag{12}$$

$$R^l_{k[j]} \;=\; R^i_j R^j_k \quad \text{and} \quad r^l_{k[j]} \;=\; r^i_j + \lambda_2 R^i_j r^j_k \tag{13}$$

Ideally, the two estimated EOPs of image k as in Equations (12) and (13) should have identical values. However, due to the uncertainty introduced in the estimated ROPs, the two EOPs are usually different. Inspired by He and Habib [81], a simple strategy, which evaluates the rotational and positional differences between the two estimated EOPs as in Equations (12) and (13), can be conducted for the compatibility analysis of all possible image triplets. Specifically, the rotational difference ($\Delta\omega$, $\Delta\phi$, $\Delta\kappa$), which describes the angular deviations between the two rotation matrices $R^l_{k[i]}$ and $R^l_{k[j]}$, can be derived through a product of $\left(R^l_{k[i]}\right)^T$ and $R^l_{k[j]}$ as shown in Equation (14). In practice, $R(\Delta\omega, \Delta\phi, \Delta\kappa)$ is expected to be close to the 3-by-3 identity matrix $I_{3\times3}$ since we usually assume small difference between the two estimated rotation matrices $R^l_{k[i]}$ and $R^l_{k[j]}$. The positional difference (Δx, Δy, Δz), which represents the discrepancies between $r^l_{k[i]}$ and $r^l_{k[j]}$, can be computed as in Equation (15). Despite the simplicity of the introduced compatibility analysis, the derived rotational and positional differences cannot be compared together as they are defined by different metrics (e.g., degrees and meters). To resolve this issue, another estimate, which quantitatively evaluates the impact of angular deviations ($\Delta\omega$, $\Delta\phi$, $\Delta\kappa$) in object space, is proposed as in Equation (16). As graphically illustrated in Figure 5, such estimate can be interpreted as the discrepancy (Δx_R, Δy_R, Δz_R) caused by the angular deviations ($\Delta\omega$, $\Delta\phi$, $\Delta\kappa$) at image k. Then, a final score function S, which considers both the rotational and positional differences in object space, can be used for the compatibility analysis (see Equation (17)). Since a high degree of similarity between the two estimated EOPs usually indicates a good geometric compatibility within the image

triplet, the second condition for initializing the local reference coordinate system can be satisfied by selecting the candidate with the minimum score value *S* in the compatibility analysis.
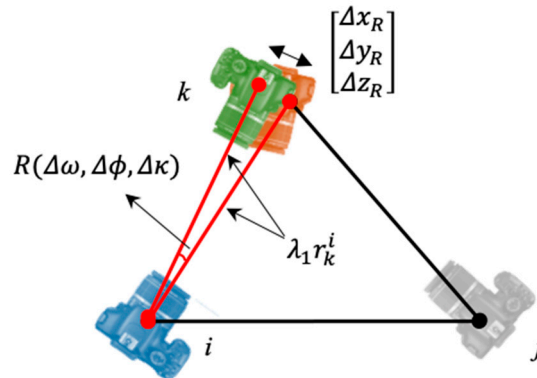

Figure 5. Discrepancy (Δx_R, Δy_R, Δz_R) caused by the angular deviations ($\Delta\omega$, $\Delta\phi$, $\Delta\kappa$) at image *k*.

$$(\Delta\omega,\ \Delta\phi,\ \Delta\kappa)\ =\ \left(R_{k[i]}^{l}\right)^{T} R_{k[j]}^{l} \tag{14}$$

$$\begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = r_{k[i]}^{l} - r_{k[j]}^{l} \tag{15}$$

$$\begin{bmatrix} \Delta x_R \\ \Delta y_R \\ \Delta z_R \end{bmatrix} = (R(\Delta\omega,\ \Delta\phi,\ \Delta\kappa) - I_{3\times 3}) \cdot \lambda_1 r_k^i \tag{16}$$

$$S\ =\ \left\| \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} \right\| + \left\| \begin{bmatrix} \Delta x_R \\ \Delta y_R \\ \Delta z_R \end{bmatrix} \right\| \text{ where, } \|\cdot\| \text{ stands for the } L_2 - \text{norm of a vector} \tag{17}$$

It is worth noting that the proposed compatibility analysis cannot handle a set of images with the linear trajectory configuration as it assumes a triangular relationship within the initial image triplet. In this case, an image stereo-pair that satisfies the requirements of a large number of corresponding points as well as a large baseline/depth ratio is selected to establish the local coordinate frame.

3.2.2. Image Augmentation Process

Once the local reference coordinate system is established, the remaining images can be sequentially augmented into the final image block or trajectory. The proposed approach for EOP recovery of each individual image is based on the tree structure introduced by Martinec and Pajdla [62]. As shown in Figure 6, the tree structure can be defined as a collection of referenced and unreferenced images. In this research, the referenced images refer to the images that have already been augmented in the local coordinate system. The unreferenced images, on the other hand, represent the remaining ones with unknown EOPs. In the given tree structure, one unreferenced image is selected as the root node; *m* (*m* > 2) referenced images are considered as leaf nodes, and possible relative orientations relating the referenced image to the unreferenced ones are used to represent the edges connecting the root and leaf nodes. In the remainder part of this section, the estimation of the rotational and positional components of the image EOPs defined by the root node is conducted separately according to the following sequence.
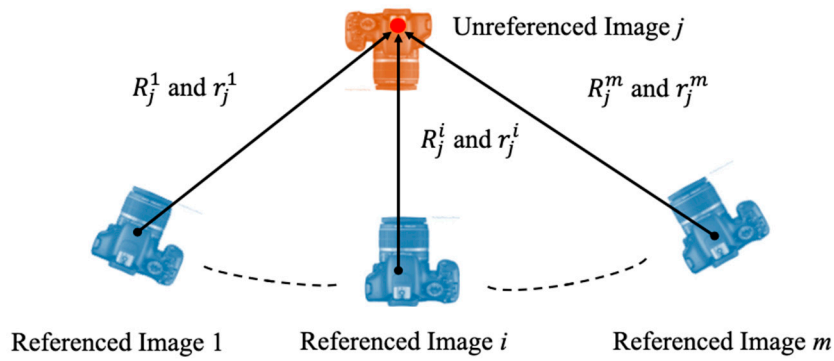
Figure 6. The tree structure for referenced (blue color) and unreferenced (red color) images.

Rotational Parameters Estimation:

Looking into Figure 6, one can establish the rotation constraint within the stereo-pair (i, j) as in Equation (18), where the rotation of the unreferenced image j (i.e., R_j^l) is expressed as a product of the rotation of the referenced image i (i.e., R_i^l) and the relative rotation R_j^i relating the two images. Given m available stereo-pairs between the referenced and unreferenced images as illustrated in the tree structure, m estimates of the rotation matrix R_j^l can be established. In this regard, the proposed approach for deriving the rotation of the unreferenced image j can be considered as a single rotation averaging problem, which aims at finding the optimum estimate of a single rotation from multiple observations [61]. In practice, averaging the m estimates of the rotation matrix R_j^l can be simply accomplished through a linear approach. Through a closer inspection of Equation (18), one can modify the presented rotation constraint to the matrix form in Equation (19), where r_{11} to r_{33} represent the nine unknown elements of R_j^l, $Y_{9 \times 1}$ is a 9-by-1 vector, and $A_{9 \times 9}$ is a 9-by-9 coefficient matrix. It is worth noting that every element in $Y_{9 \times 1}$ and $A_{9 \times 9}$ is defined by a numeric value. According to Equation (19), a system of $9m$ linear equations can be established for the m possible stereo-pairs within the tree structure. Then, a least-squares solution for the nine unknown elements (i.e., r_{11} to r_{33}) can be derived. An alternative linear solution for the single rotation averaging is achieved through quaternion [62]. Compared to the rotation matrix approach, the utilization of quaternion can be more advantageous as it only requires four elements for the representation of a rotation. As a result, the total number derived linear equations is reduced from $9m$ to $4m$. However, one has to note that both the rotation matrix and the quaternion-based approaches fail to consider the inherent constraint within a rotation (i.e., the orthogonal constraints for the rotation matrix and the unit length constraint for the quaternion).

$$R_j^l = R_i^l R_j^i \Rightarrow R_j^i = \left(R_i^l\right)^T R_j^l \text{ where, } R_j^l = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{23} & r_{33} \end{bmatrix} \tag{18}$$

$$Y_{9 \times 1} = A_{9 \times 9} \begin{bmatrix} r_{11} \\ r_{12} \\ r_{13} \\ r_{21} \\ r_{22} \\ r_{23} \\ r_{31} \\ r_{32} \\ r_{33} \end{bmatrix} \tag{19}$$

In this research, a new quaternion-based approach has been proposed for single rotation averaging while enforcing the unit length constraint on the resulting quaternion. The conceptual basis of the proposed approach is the quaternion-based solution, which is initially proposed by Horn [82], and further investigated by Guan and Zhang [83] and He and Habib [84]. This solution is originally designed for the recovery of absolute orientation parameters between two datasets while using two sets of conjugate vectors with compatible directions in these datasets. However, our proposed approach is the first attempt to adopt such concept for resolving the single rotation averaging problem. As can be seen in Equation 18, the rotation matrix R_j^l describes the rotation between the camera coordinate system of the image j and the local coordinate system l. Therefore, the proposed approach starts with generating two sets of conjugate vectors in the two coordinate systems. In this research, the three unit vectors along the x, y, and z axes in the camera coordinate system of the image j are first selected. Then, the derived rotation matrix R_j^l of each possible stereo-pair (see Equation (18)) is applied to the three vectors to convert them into the the local coordinate system l. Given m available stereo-pairs in the tree structure, $3m$ pairs of conjugate unit vectors can be established. For each pair of conjugate vectors, one can introduce a parallelism constraint as in Equation (20).

$$\vec{v}_{[d]}^{\,l\{i\}} \;=\; R_j^l\,\vec{v}_{[d]} \;=\; R_i^l R_j^i\,\vec{v}_{[d]} \quad (i = 1, \cdots, m \ \text{ and } \ d = x, y, \text{ or } z) \tag{20}$$

In Equation (20), $\vec{v}_{[d]}$ and $\vec{v}_{[d]}^{\,l\{i\}}$ are two conjugate unit vectors defined in the camera coordinate system of image j and the local coordinate system l, respectively. In $\vec{v}_{[d]}$ and $\vec{v}_{[d]}^{\,l\{i\}}$, the subscript $[d]$ ($d = x, y, \text{ or } z$) describes the direction of the unit vector (i.e., x, y, or z axes of the coordinate system), and the superscript $\{i\}$ indicates the utilized stereo-pair (i, j) for establishing the parallelism constraint. For the $3m$ pairs of conjugate unit vectors, a set of $3m$ the equations as the form in Equation (20) can be established. To derive the optimum estimate of the rotation matrix R_j^l, a least-squares approach is adopted to minimize the sum of squared errors (SSE) for all the involved 3m pairs of conjugate unit vectors (see Equation (21)). In Equation (21), the terms $\left(\vec{v}_{[d]}^{\,l\{i\}}\right)^T \vec{v}_{[d]}^{\,l\{i\}}$ and $\left(\vec{v}_{[d]}\right)^T \vec{v}_{[d]}$ are always equal to 1 as they are the squared magnitudes of the unit vectors $\vec{v}_{[d]}^{\,l\{i\}}$ and $\vec{v}_{[d]}$. Therefore, to minimize the SSE, the rotation matrix R_j^l has to be estimated in such a way to maximize to the term $\left(\vec{v}_{[d]}^{\,l\{i\}}\right)^T R_j^l\,\vec{v}_{[d]}$. One should note that the term $\left(\vec{v}_{[d]}^{\,l\{i\}}\right)^T R_j^l\,\vec{v}_{[d]}$ is always positive as $R_j^l\,\vec{v}_{[d]}$ and $\vec{v}_{[d]}^{\,l\{i\}}$ are always pointing in the same direction. Then, this term can be formulated as a dot product as in Equation (22). According to quaternion properties, the rotation multiplication $R_j^l\,\vec{v}_{[d]}$ is equivalent to the quaternion multiplication $\dot{q}\dot{v}_{[d]}\dot{q}^*$, where the unit quaternion $\dot{q}$ corresponds to R_j^l, and $\dot{q}^*$ is the conjugate quaternion constructed by negating the imaginary part of $\dot{q}$. The term $\dot{v}_{[d]}$ is the quaternion form of $\vec{v}_{[d]}$, which is simply adding a zero as the real part and the three elements of $\vec{v}_{[d]}$ as the imaginary part. Using the quaternion properties, Equation (22) can be rewritten as in Equation (23), where C and $\bar{\bar{C}}$ are 4-by-4 matrices that convert the quaternion-based multiplication to a matrix-based multiplication, and the summation matrix S is a 4-by-4 matrix constructed using the components of $\vec{v}_{[d]}^{\,l\{i\}}$ and $\vec{v}_{[d]}$

for all the conjugate vector pairs.

$$\min_{R_j^l} \sum_{i=1}^{m} \sum_{d\in \{x,y,z\}} (\overrightarrow{v}_{[d]}^{l\{i\}} - R_j^l \overrightarrow{v}_{[d]})^T \left(\overrightarrow{v}_{[d]}^{l\{i\}} - R_j^l \overrightarrow{v}_{[d]} \right)$$
$$= \min_{R_j^l} \sum_{i=1}^{m} \sum_{d\in \{x,y,z\}} \left((\overrightarrow{v}_{[d]}^{l\{i\}})^T \overrightarrow{v}_{[d]}^{l\{i\}} + (\overrightarrow{v}_{[d]})^T \overrightarrow{v}_{[d]} - 2 \left(\overrightarrow{v}_{[d]}^{l\{i\}} \right)^T R_j^l \overrightarrow{v}_{[d]} \right) \tag{21}$$

$$\max_{R_j^l} \sum_{i=1}^{m} \sum_{d\in \{x,y,z\}} (\overrightarrow{v}_{[d]}^{l\{i\}})^T R_j^l \overrightarrow{v}_{[d]} = \max_{R_j^l} \sum_{i=1}^{n} \sum_{d\in \{x,y,z\}} \overrightarrow{v}_{[d]}^{l\{i\}} \cdot \left(R_j^l \overrightarrow{v}_{[d]} \right) \tag{22}$$

$$\max_{\dot{q}} \sum_{i=1}^{m} \sum_{d\in \{x,y,z\}} \dot{v}_{[d]}^{l\{i\}} \cdot \dot{q} \dot{v}_{[d]} \dot{q}^* = \max_{\dot{q}} \sum_{i=1}^{m} \sum_{d\in \{x,y,z\}} \left(\dot{v}_{[d]}^{l\{i\}} \dot{q} \right) \cdot \left(\dot{q} \dot{v}_{[d]} \right)$$
$$= \max_{\dot{q}} \sum_{i=1}^{m} \sum_{d\in \{x,y,z\}} \left(C\left(\dot{v}_{[d]}^{l\{i\}} \right) \dot{q} \right) \cdot \left(\overline{C}\left(\dot{v}_{[d]} \right) \dot{q} \right)$$
$$= \max_{\dot{q}} \sum_{i=1}^{m} \sum_{d\in \{x,y,z\}} \dot{q}^T C(\dot{v}_{[d]}^{l\{i\}})^T \overline{C}\left(\dot{v}_{[d]} \right) \dot{q} \tag{23}$$
$$= \max_{\dot{q}} \dot{q}^T \left(\sum_{i=1}^{m} \sum_{d\in \{x,y,z\}} C(\dot{v}_{[d]}^{l\{i\}})^T \overline{C}\left(\dot{v}_{[d]} \right) \right) \dot{q}$$
$$= \max_{\dot{q}} \dot{q}^T S \dot{q}$$

To maximizes the term $\dot{q}^T S \dot{q}$ while maintaining the unity constraint of $\dot{q}$ (see Equation (24)), one has to maximize the target function φ using the Lagrange multiplier λ as in Equation (25). The partial derivative of the target function φ with respected to $\dot{q}$ is then established as the expression in Equation (26). Equation (26) is satisfied if and only if λ and $\dot{q}$ are the corresponding eigenvalues and eigenvectors of the summation matrix S. Given the fact that $\dot{q}^T S \dot{q}$ is maximized when λ is the largest eigenvalue of S, the quaternion $\dot{q}$ is eventually the eigenvector corresponding to this largest eigenvalue. Finally, the rotation matrix R_j^l is recovered through the estimated quaternion $\dot{q}$.

It is worth noting that the proposed solution for R_j^l can be integrated within a RANSAC framework for outlier removal. Specifically, a single stereo-pair is first randomly selected between the unreferenced image j and the set of referenced images within the involved tree structure. Then, the ROPs of this stereo-pair is used to derive an estimate of the rotation matrix R_j^l as presented in Equation (18). Afterwards, the rotational errors are evaluated for all the remaining stereo-pairs according to the rotation constraint $(R_j^l)^T \left(R_i^l R_j^i \right) = I_{3\times3}$. Such a sampling-and-testing procedure is repeated until the random selection resulting in the largest consensus is achieved. All these inlier stereo-pairs are finally used in the proposed single rotation averaging to derive a reliable estimate of the rotation matrix R_j^l.

$$\max_{\dot{q}} \dot{q}^T S \dot{q}, \quad \|\dot{q}\| = 1 \tag{24}$$

$$\max_{\dot{q}} \varphi(\dot{q}) = \dot{q}^T S \dot{q} - \lambda \left(\dot{q}^T \dot{q} - 1 \right) \tag{25}$$

$$\frac{\partial \varphi}{\partial \dot{q}} = 2S\dot{q} - 2\lambda \dot{q} = 0 \quad \Rightarrow \quad S\dot{q} = \lambda \dot{q} \tag{26}$$

Positional Parameters Estimation:

The positional parameters of the unreferenced image j is separately estimated relative to the local reference coordinate system through two different closed-form solutions. These two solutions are designed to handle UAV images that are captured either in a block or a linear trajectory configuration, respectively. As for the images within a block configuration, the positional parameters are derived through an intersection of multiple vectors. As shown in Figure 6, these vectors are the translations connecting the referenced images to the unreferenced one within the tree structure. The mathematical model for the multi-vector intersection is presented in Equation (27), where r_j^l stands for the positional parameters of the unreferenced image j defined in the local coordinate system.

$$r_j^l = r_i^l + \lambda_i R_i^l r_j^i \tag{27}$$

Looking into Equation (27), one can note that r_j^l is expressed as a summation of two vectors: r_i^l and $\lambda_i R_i^l r_j^i$. Specifically, r_i^l is the position of the referenced image i defined in the local coordinate system while $\lambda_i R_i^l r_j^i$ represents the translation relating the two images. The vector $\lambda_i R_i^l r_j^i$ can be derived in two steps. First, the rotation matrix R_i^l is applied to the relative translation r_j^i to convert it to the local coordinate system l. Then, considering the fact that the relative translation r_j^i is defined with an arbitrary scale, a scale factor λ_i is multiplied to $R_i^l r_j^i$ to get the translation vector with correct scale between the two involved images. Assuming m available stereo-pairs connecting the set referenced image to the unreferenced one, one would have a system of $3m$ equations in $m + 3$ unknowns (i.e., m unknown scale factors λ_i and three unknown parameters for r_j^l). In this regard, a minimum of two intersecting translation vectors (i.e., two stereo-pairs) would allow for a least-squares solution for the positional parameters r_j^l as well as the unknown scale factors λ_i.

It is worth noting that the mathematical model as presented in Equation (27) assumes non-collinear relationship among multiple vectors. Given a set of UAV images captured in a linear trajectory configuration (i.e., all images are acquired in a single straight flight path), the proposed multi-vector intersection model leads to a degenerate case, in which all involved translation vectors are collinearly aligned. In order to derive a reliable estimate of the positional parameters for these linear trajectory images, we first conduct feature tracking among the referenced and unreferenced images according to the introduced tree structure. In the feature tracking process, a set of corresponding features among overlapping images is called a feature track. In this research, the derived feature tracks within the tree structure should include corresponding points visible in the unreferenced image j as well as all/some of the referenced images. In addition, the minimum length of the accepted feature tracks should be greater or equal to three, which, in other words, means that the tracked tie points should be visible in at least three images. Now that the feature correspondences are established, the 3D object coordinates of these tracked points can be derived through a spatial intersection while using the EOPs of the referenced images. As illustrated in Figure 7, these reconstructed object points can be then utilized to recover the translation parameters of the unreferenced image. The mathematical expression for deriving the positional parameters of the unreferenced image j, which have been denoted as r_j^l, is shown in Equation (28).

$$r_j^l = s_i R_j^l p_i^j + \begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix} \qquad \text{where,} \quad p_i^j = \left(x_i^j, \ y_i^j, \ -c^j \right)^T \tag{28}$$

In Equation (28), $(X_i, \ Y_i, \ Z_i)^T$ stands for the 3D coordinates of the object point P_i $(i = 1, \cdots, m)$. The vector $p_i^j = \left(x_i^j, \ y_i^j, \ -c^j \right)^T$ represents the image coordinates of P_i in the unreferenced image j

after correcting for the principal point offset and camera-specific distortions. The rotation matrix R_j^l is derived through the presented single rotation averaging approach, and converts the vector of image coordinates to the local coordinate system. s_i is the unknown scale factor for the vector $R_i^l p_i^j$. It is worth noting that, in this paper, we always denote the scale factor for the translations among overlapping images as λ while using s for the image vector connecting the object point to the image. Given m object points, a system of $3m$ equations in in $m + 3$ unknowns can be established according to Equation (28). Although a minimum of two object points is sufficient for solving the unknown positional parameters r_j^l, redundant points with good spatial distribution are always preferred to achieve a more reliable estimation. One has to note that different from the existing single photo resection approaches, which directly recovers the EOPs of the involved images through the coordinates of object points, the closed-form solution as presented in Equation (28) only recovers the positional parameters while the rotation of the image is separately estimated through the introduced single rotation averaging.
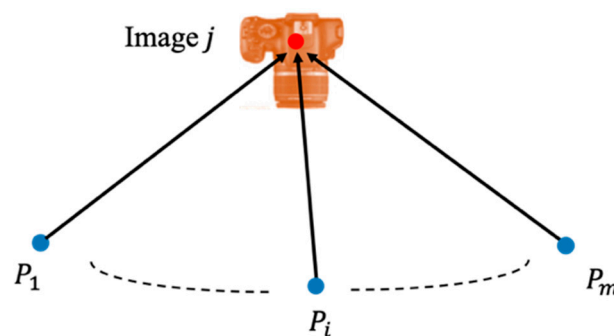
Figure 7. Estimation of the positional parameters of image *j* through multiple object points.

Compatibility Analysis for Image Augmentation:

Similar to most existing incremental approaches, the proposed image augmentation for the initial recovery of image EOPs suffers from the accumulated drifting errors. In order to mitigate the impact of error propagation, the proposed incremental strategy is based on augmenting images that exhibit the best compatibility with previously referenced images to establish the final image block or trajectory. In this research, for a set of previously referenced images, we check all possible unreferenced images that could be augmented. More specifically, the overall residuals derived from the rotational/positional parameters estimation are used to evaluate the compatibility among the referenced and unreferenced images within the tree structure. Only the image that exhibits the highest compatibility (i.e., lowest residuals) with the set of previously referenced imagery is selected and referenced into the current image network at each step of the image incremental augmentation.

3.3. Global Strategy for EOP Recovery

Apart from the incremental strategy, a global approach, which simultaneously establishes the rotational and positional estimation of all involved images, has been investigated in this research. The proposed global approach is implemented in two steps. In the first step, a multiple rotation averaging is conducted for deriving the rotational parameters of all involved imagery. Then, the positional parameters are determined through a global translation averaging while using the derived rotations in the first step.

3.3.1. Multiple Rotation Averaging

Given a set of estimated ROPs from the relative orientation procedure, the multiple rotation averaging aims at providing a direct rotation estimation for all involved imagery. In this research, a simple global rotation averaging approach, which is based on the rotation constraint as presented

in Equation (18), is adopted. Different from the single rotation averaging, in which only the rotation matrix R_j^l has to be estimated, both R_i^l and R_j^l are unknown in the proposed multiple rotation averaging. For a single stereo-pair, the rotation constraint can be represented using the matrix form in Equation (29), where $A_{9 \times 18}$ is a 9-by-18 coefficient matrix (i.e., each element in the matrix is defined by a numeric value). 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 in a matrix form of Equation (30), where $A_{9m \times 9n}$ is the $9m$-by-$9n$ coefficient matrix, and $X_{9n \times 1}$ is the $9n$-by-1 vector containing all the elements in the unknown rotation matrices. A closed-form solution for the unknown vector $X_{9n \times 1}$ can be derived through the eigenvector corresponding to the smallest eigenvalue in $(A_{9m \times 9n})^T A_{9m \times 9n}$. Unfortunately, such a solution ignores the inherent orthogonality constraint within the estimated rotation matrix. In practice, more accurate estimate of the rotational parameters can be achieved through a non-linear iterative refinement while enforcing the orthogonal constraints among the elements of the rotation matrices.

$$R_j^l = R_i^l R_j^i \quad \Rightarrow \quad A_{9 \times 18} \begin{bmatrix} r_{11}^{\{i\}} \\ \vdots \\ r_{33}^{\{i\}} \\ r_{11}^{\{j\}} \\ \vdots \\ r_{33}^{\{j\}} \end{bmatrix} = \mathbf{0}_{18 \times 1} \tag{29}$$

$$\text{where, } R_i^l = \begin{bmatrix} r_{11}^{\{i\}} & r_{12}^{\{i\}} & r_{13}^{\{i\}} \\ r_{21}^{\{i\}} & r_{22}^{\{i\}} & r_{23}^{\{i\}} \\ r_{31}^{\{i\}} & r_{32}^{\{i\}} & r_{33}^{\{i\}} \end{bmatrix} \text{ and } R_j^l = \begin{bmatrix} r_{11}^{\{j\}} & r_{12}^{\{j\}} & r_{13}^{\{j\}} \\ r_{21}^{\{j\}} & r_{22}^{\{j\}} & r_{23}^{\{j\}} \\ r_{31}^{\{j\}} & r_{32}^{\{j\}} & r_{33}^{\{j\}} \end{bmatrix}$$

$$A_{9m \times 9n} X_{9n \times 1} = \mathbf{0}_{9m \times 1} \tag{30}$$

3.3.2. Global Translation Averaging

The proposed global translation averaging starts with generating a graph structure for the set of involved images. As illustrated in Figure 8a, each node in the graph represents one image, and each edge connecting two nodes indicates an available relative orientation between two overlapping images. To estimate the positional parameters, the graph structure can be further divided into several sub-graphs. In Figure 8b, the sub-graph established on image j, which is similar to the tree structure shown in Figure 6, includes one root node (i.e., image j), a few leaf nodes (i.e., all images connected to image j), and edges connecting the root and leaf nodes (i.e., stereo-based relative orientation). Based on such graph and sub-graph structures, two different types of constraints can be established for estimating the positional parameters of each involved image. The first type of constraints—namely the translation constraint—is based on the same mathematical expression as presented in Equation (27), which describes the fact that within the stereo-pair (i, j), the positional parameters of image j (i.e., r_j^l) can be defined by a summation of the position of image i (i.e., r_i^l) and the scaled translation vector between the two images (i.e., $\lambda_i R_i^l r_j^i$). The second type of constraints is based on the conjugate point pairs among overlapping images. As can be seen in Figure 9, the conjugate point constraint can be formulated as an intersection of two image vectors. Such a vector intersection can be formulated as in Equation (31), where $p_i = (x_i, y_i, -c_i)$ and $p_j = (x_j, y_j, -c_j)$ are the image coordinates for two conjugate points after correcting for the principal point offsets and camera-specific distortions, and s_i and s_j are the two scale factors for p_i and p_j, respectively.

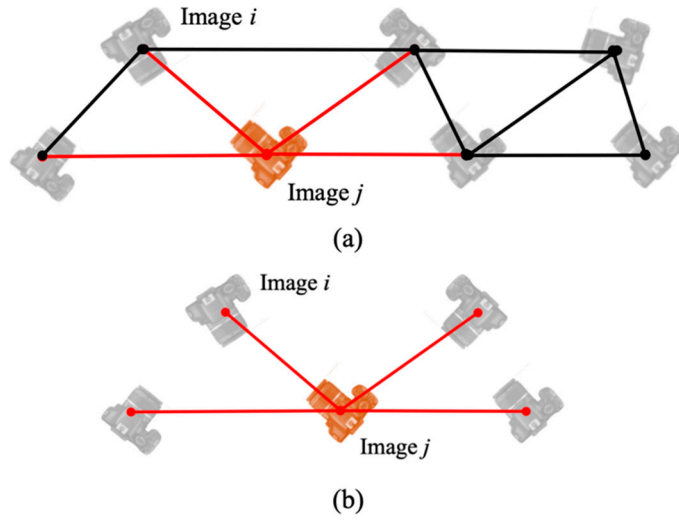$$r_i^l + s_i R_i^l p_i = r_j^l + s_j R_j^l p_j \tag{31}$$

Figure 8. (**a**) The graph structure for a block of overlapping images, and (**b**) the established sub-graph at image j.
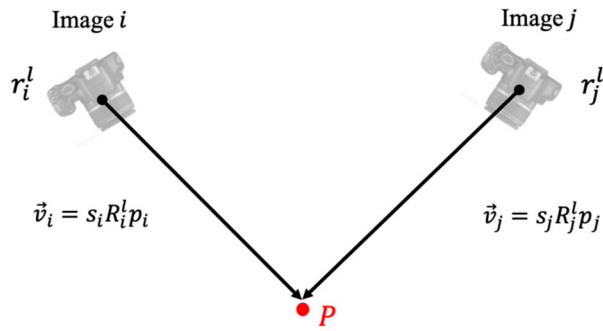

Figure 9. The established conjugate point constraint for global translation averaging.

In practice, given a set of n overlapping images with m available stereo-pairs, m translation-based constraints can be established through Equation (27). Meanwhile, suppose that a total number of o pairs of conjugate points can be identified within the given image dataset, o conjugate point constraints can be formulated as in Equation (31). It is worth noting that there is a total number of $(3n + m + 2o)$ unknown parameters within the $(3m + 3o)$ derived equations (i.e., each translation/conjugate-point constraint leads to three linear equations). To derive a solution for the $(3n + m + 2o)$ unknown parameters, the condition $3m + 3o \geq 3n + m + 2o$ has to be satisfied. The $(3n + m + 2o)$ unknown parameters include: $3n$ unknown positional parameters (i.e., one image has 3 unknown positional parameters), and $m + 2o$ unknown scale factors (i.e., each translation constraint provides one unknown scale factor λ_i, and every conjugate point constraint gives two unknown scale factors s_i and s_j). A matrix form for the set of established constraints can be established as in Equation (32), where $A_{(3m+3o) \times (3n+m+2o)}$ is a coefficient matrix, and $X_{(3n+m+2o) \times 1}$ is a vector comprised of all the unknown parameters. Similar to the proposed multiple rotation averaging, the closed-form solution for the unknown vector $X_{(3n+m+2o) \times 1}$ corresponds to the eigenvector of the smallest eigenvalue in $\left(A_{(3m+3o) \times (3n+m+2o)}\right)^T A_{(3m+3o) \times (3n+m+2o)}$. Although all identified conjugate points in stereo-pairs can be used for the proposed global translation averaging, the excessive number of unknown parameters regarding the scale factors s_i and s_j leads to a huge and sparse matrix of A. Such huge matrix size makes the computation really expensive and even impossible. In order to

resolve this problem, the sparse matrix technique is utilized for the representation of matrix in this research. In addition, to reduce the total number of unknown parameters, the proposed approach only randomly selects 10 pairs of conjugate points from each stereo-pair. Compared to the introduced solution for the incremental strategy, the proposed global translation averaging is capable of dealing with images captured either in a block or a linear trajectory configuration due to the utilization of both translation and conjugate point constraints.

$$A_{(3m+3o)\times(3n+m+2o)} X_{(3n+m+2o)\times 1} = 0 \tag{32}$$

3.4. Global Bundle Adjustment

It is worth noting that the proposed incremental and global approaches for the EOP recovery are both conducted in an arbitrary local coordinate system. In order to geo-reference the derived 3D model, an initial 3D Helmert transformation [85] is required to establish the transformation from the local coordinate system to the mapping frame. Such initial 3D Helmert transformation provides approximations for further bundle adjustment of the estimated parameters in the mapping coordinate system. In this research, to estimate the 3D Helmert transformation parameters (i.e., scale factor, three translation parameters, and three rotation angles) relating local and mapping coordinate systems, tie points corresponding to the GCPs are first manually identified. Then, the 3D coordinates of tie points, which are defined in the local coordinate system, are computed through a spatial intersection. Afterwards, the transformation parameters are estimated using the GCPs and their corresponding local coordinates. Finally, the set of estimated transformation parameters are used to convert the estimated image EOPs as well as 3D object coordinates from the local coordinate system to the mapping frame. After conducting the initial 3D Helmert transformation, a global bundle adjustment is adopted to refine the geo-referenced image EOPs and object coordinates as well as the GPS-surveyed ground control/check points. In the implemented global bundle adjustment, GCPs are established for absolute orientation/datum definition. The inputs and outputs for the global bundle adjustment process are illustrated in Figure 10.
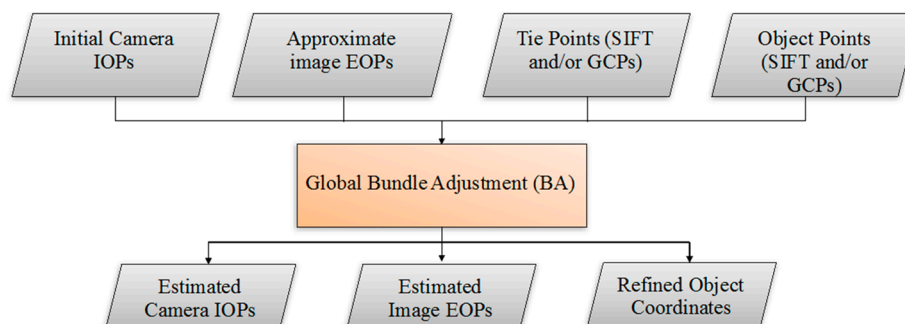

Figure 10. The inputs and outputs for the adopted global bundle adjustment.

4. Experimental Results

The main objective of the experimental results is illustrating the feasibility of the proposed framework for automated aerial triangulation while using UAV-based imagery with different "texture" characteristics. In this research, the concept of "texture" refers to the number of unique image features that can be identified within image stereo-pairs. More specifically, images with sufficient or strong "texture" always lead to a large number of unique features that can be used to robustly estimate ROPs within stereo-pairs, while only very few conjugate point pairs can be identified within stereo-pairs with poor "texture".

4.1. Data Description

Two different types of test sites are involved for the experimental tests. The first test site covers agriculture fields with repetitive patterns. Such repetitive patterns might introduce point features with significant ambiguities in the image matching procedure. The second one is at the vicinity of a building with a complex roof structure, and the texture on the building rooftop is capable of providing a large number of unique point features for image matching and relative orientation recovery. In this research, three datasets arising from the agriculture field, and one dataset captured for the building have been acquired by two different UAV platforms. The two utilized UAVs include: a DJI Phantom 2 UAV with a GoPro Hero 3+ black edition camera (see Figure 11a), and a DJI S1000+ UAV with a Sony Alpha 7R camera (see Figure 11b). The specifications of the utilized UAVs and cameras are reported in Table 1. The internal characteristics of the utilized cameras are estimated through a calibration procedure similar to the one proposed by He and Habib [86], where the USGS Simultaneous Multi-frame Analytical Calibration (SMAC) distortion model is adopted. For years, the USGS has been using the SMAC model for calibrating both film and digital cameras. In the SMAC model, all image points must be referenced to the center of image coordinate system with the correct principal distance c and the principal point (x_p, y_p). Regarding the distortion parameters, the model considers both radial and de-centering lens distortions. In this research, due to the significant image distortions caused by the wide-angle lens, three radial lens distortion parameters K_1, K_2, and K_3 and two de-centering lens distortion parameters P_1 and P_2 are used for the calibration of the GoPro Hero 3+ black edition camera. On the other hand, only four distortion parameters (K_1, K_2, P_1, P_2) are considered for the Sony Alpha 7R camera. One should note that for the DJI Phantom 2 UAV, the GoPro camera is mounted on a gimbal to ensure that images are acquired with the camera's optical axis pointing in the nadir direction. On the other hand, the Sony Alpha 7R camera is rigidly fixed to the body of the DJI S1000+ UAV platform while pointing in the nadir direction. According to such configuration of the cameras, images captured by the utilized UAVs can be assumed to comply with the assumption of the two-point approach (see Section 3.1) for relative orientation recovery. The details pertaining to the four utilized image datasets for the experimental tests are provided in the following paragraph.

DJI Phantom 2 & GoPro Hero 3+

(a)

DJI S1000+ & Sony Alpha 7R Camera

(b)

Figure 11. (**a**) The DJI Phantom 2 UAV with a GoPro Hero 3+ camera mounted on a gimbal, and (**b**) the DJI S1000+ UAV equipped with a Sony Alpha 7R camera with a vertical view.

Table 1. Specification of the utilized UAV and cameras.

Specs/Model	UAVs	
	DJI Phantom 2	**DJI S1000+**
Weight	1000 g (Take-off weight: < 1300 g)	4.2 kg (Take-off weight: 6.0 kg ~ 11.0 kg)
Max Speed	15 m/s	20 m/s
Max Flight Endurance	Approximate 23 min	15 min (9.5 kg take-off weight)
Diagonal Size	350 mm	1045 mm
Specs/Model	Cameras	
	GoPro Hero 3+ Black Edition	**Sony Alpha 7R**
Image Size	3000 × 2250 pixels (medium field-of-view)	7360 × 4912 pixels
Pixel Size	1.55 μm	4.90 μm
Focal Length	3.5 mm	35 mm

Phantom2-Agricuulture Dataset is comprised of 569 images that are captured from a flying height of 15 m while moving at a speed of roughly 8 m/s. The overlap and side lap percentages for the acquired images are approximately 60%. The Ground Sampling Distance (GSD) is about 0.7 cm.

Phantom2-Building Dataset includes 81 images acquired with a speed of roughly 4 m/s from a flying height of roughly 20 m. The overlap and side lap percentages for the acquired images are approximately 80% and 60%, respectively. The GSD is about 0.9 cm.

S1000-Agriculture-1 Dataset includes 421 images that are captured from a flying height of 50 m with a flying speed of roughly 5 m/s. The overlap and side lap percentage of the acquired images are approximately 78% and 73%, respectively. The GSD is about 0.7 cm.

S1000-Agriculture-2 Dataset has 639 images, which are captured by the DJI S1000+ UAV while flying at a speed of 8 m/s at a flying height of almost 40 m. The overlap and side lap percentages for the acquired images are approximate 70%. The GSD is about 0.6 cm.

Figure 12 shows sample images that are captured with different characteristics from the four experimental datasets. Figure 12a illustrate the image captured by the GoPro Hero 3+ camera above one agriculture field with repetitive patterns. Due to the large field-of-view (FOV) of the utilized GoPro camera, one can observe significant image distortions in the acquired image. On the other hand, Figure 12b exhibits the image, which was acquired by the same GoPro camera, over building roof-top with rich "texture". The rich "texture" in the acquired image (see Figure 12b) leads to an adequate number of features for image matching. Figure 12c presents the image that was taken by the Sony Alpha 7R camera before planting any crop in the field. The bare ground as shown in the image only provides a limited number of identified features. The sample image in Figure 12d captured by the same Sony camera shows the repetitive pattern over mature crops within the agricultural test field.
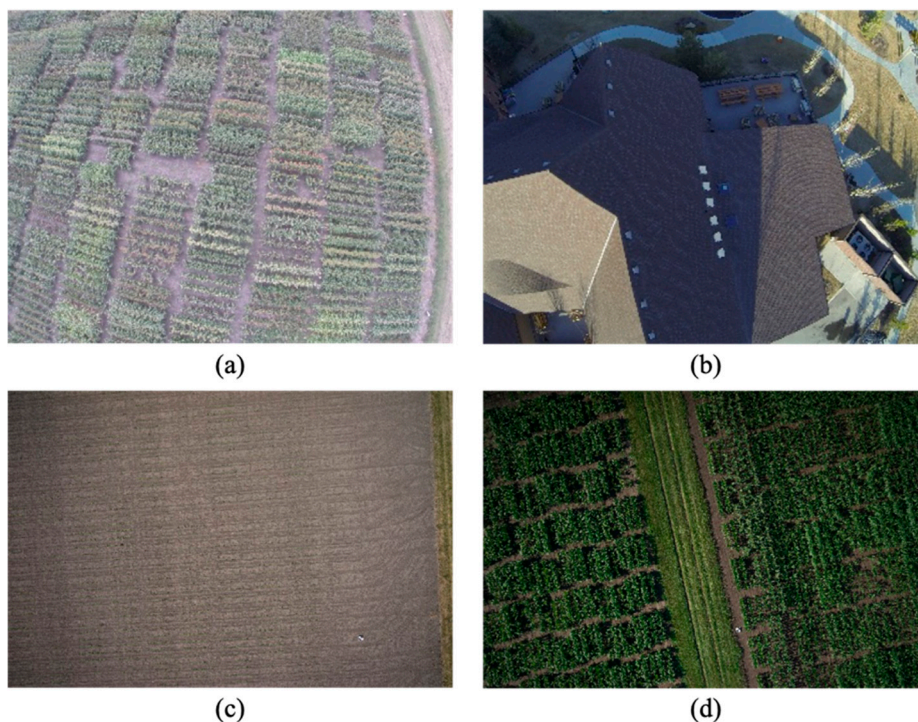
Figure 12. (**a**) Sample images from (**a**) Phantom2-Agriculture, (**b**) Phantom2-Building, (**c**) S1000-Agriculture-1, and (**d**) S1000-Agriculture-2 Datasets.

4.2. Comparison between Incremental and Global Estimation for Image EOPs

The objective of the first stage of the experimental results is a comparative analysis between the proposed incremental and global approaches for the initial recovery of image EOPs. In this research, such comparative analysis is performed through a quantitative comparison between the incremental/global image EOPs and those derived from the global bundle adjustment refinement (i.e., BA-based EOPs). In this comparison, the same set of inlier stereo-pairs, which are identified in the incremental approach (the incremental approach is coupled with a built-in process for outlier detection/removal), are utilized in the global approach for the initial recovery of image EOPs. In addition, since the two sets of estimated EOPs (i.e., incremental, global) are defined in different local coordinate systems, a 3D Helmert transformation has to be conducted to transform the derived incremental/global parameters as well as the BA-based EOPs to a common reference coordinate system for the comparison. To estimate the transformation parameters converting the estimated incremental/global parameters to the reference frame, the mathematical formula as in Equation (33) can be used. In this equation, R_{BA} and r_{BA} represent the image rotational and positional parameters that are derived in the bundle adjustment. $R_{incremental/global}$ and $r_{incremental/global}$ stand for the corresponding orientations and positions estimated through either the incremental or global approach. s, R, and t are the 3D Helmert transformation parameters relating the estimated incremental/global EOPs to the BA-based EOPs. Given two sets of corresponding image EOPs defined in different coordinate systems, an estimate of the 3D Helmert transformation parameters can be derived through the introduced closed-form solution by Horn [82]. It is worth noting that different from the initial 3D Helmert transformation for the geo-referencing of the derived 3D model (see Section 3.4), the transformation as shown in Equation (33) is only used to compute the RMSE values between the incremental/global and BA-based EOPs.

$$r_{BA} = s \cdot R r_{incremental/global} + t \quad \text{and} \quad R_{BA} = R R_{incremental/global} \tag{33}$$

Table 2 presents the statistics of the derived differences between the transformed incremental/global and BA-based EOPs. Specifically, in Table 2, rows 1 through 6 present the derived root-mean-square error (RMSE) differences for both rotational and positional parameters when comparing the transformed incremental/global EOPs to the BA-based ones. Row 7 illustrates the processing time for the proposed incremental and global approaches in the four experimental datasets. Through a closer inspection of the reported results, one can conclude the following:

- Compared to the incremental approach, the global approach provides more accurate initial estimates of image EOPs when compared to the BA-based EOPs.
- According to the reported processing time, one can note that the global approach is more efficient when dealing with datasets including a large number of images.

Table 2. Comparison between the incremental/global and the bundle adjustment (BA)-based exterior orientation parameters (EOPs).

	Datasets							
Values	**Phantom2-Agriculture**		**Phantom2-Building**		**S1000-Agriculture-1**		**S1000-Agriculture-2**	
	Incremental	**Global**	**Incremental**	**Global**	**Incremental**	**Global**	**Incremental**	**Global**
$RMSE_\omega(°)$	1.12	0.83	0.61	0.43	0.29	0.17	1.52	1.24
$RMSE_\phi(°)$	1.79	0.77	0.88	0.65	0.55	0.31	1.71	1.63
$RMSE_\kappa(°)$	0.94	0.46	0.53	0.37	0.27	0.14	1.86	1.49
$RMSE_{Xo}(m)$	0.44	0.28	0.28	0.21	0.84	0.40	0.80	0.46
$RMSE_{Yo}(m)$	0.53	0.33	0.25	0.20	0.69	0.37	0.96	0.51
$RMSE_{Zo}(m)$	0.78	0.43	0.38	0.24	0.79	0.55	0.85	0.60
Time (min)	92.8	22.1	20.1	5.8	43.7	11.2	137.2	31.3

4.3. Accuracy Analysis

At the second stage of the experimental results, the accuracy of the derived 3D model is evaluated for each experimental dataset. More specifically, the reconstruction accuracy of **Phantom2-Agriculture**, **S1000-Agriculture-1**, and **S1000-Agriculture-2** Datasets are evaluated through check point analysis. Due to the absence of ground control information, a LiDAR point cloud, which was acquired by an Optech ALTM 3100 airborne laser scanning system, is used for the accuracy analysis of **Phantom2-Building** Dataset.

4.3.1. Check Point Analysis

In order to evaluate the accuracy of the derived 3D point clouds, both GCPs and check points, which are established on the signalized targets in the field, are surveyed by an RTK GPS with an approximate accuracy of 2 cm, and utilized in the proposed automated aerial triangulation for **Phantom2-Agriculture**, **S1000-Agriculture-1**, and **S1000-Agriculture-2** Datasets. Figure 13a–c shows the configuration of the utilized GCPs and check points for the three experimental datasets, and Figure 13d present a sample image for the utilized target. As shown in Section 3.4, an initial 3D Helmert transformation, which is followed by a global bundle adjustment procedure using GCPs and check points, is conducted at the final stage of the proposed framework to refine all the derived parameters. Once the global bundle adjustment is completed, the derived square root of a-posteriori variance factor σ_o can be computed, and the RMSE values for the check points can be evaluated (the GCPs are totally fixed in the global bundle adjustment). Table 3 reports the number of utilized GCPs and check points, derived σ_o, check point RMSE values, and the extent of covered area for the three experimental datasets. Looking into Table 3, one can observe that the derived σ_o values for the three experimental datasets are all smaller than 1.5 pixels, which indicate a good precision of the conducted bundle adjustment. In terms of the RMSE analysis of the check points, one can note that the derived RMSE values for both planimetric (i.e., X and Y) and vertical (i.e., Z) coordinates are

below 0.04 m. Such results indicate a good accuracy of the derived 3D reconstruction in object space. It is worth noting that due to the limited access to the test sites, only S1000-Agriculture-2 Dataset has check points established in middle of the involved agricultural field. The utilization of these check points leads to the larger RMSE values of **S1000-Agriculture-2** Dataset when compared to the other two. Such large RMSE values would be expected in both **Phantom2-Agriculture**, and **S1000-Agriculture-2** Datasets if additional check points are available in the middle of the agricultural field.
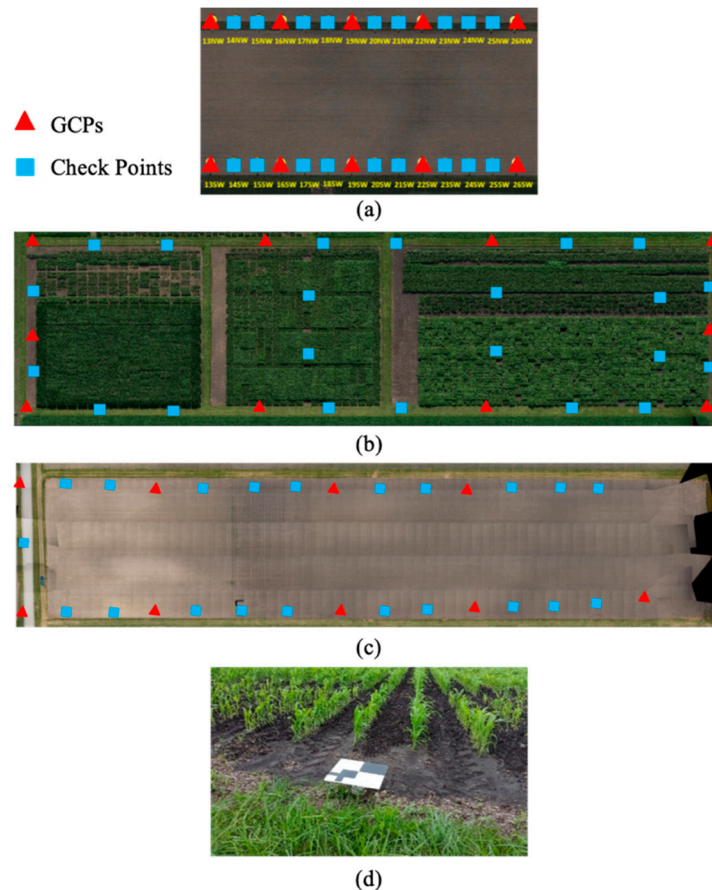
Figure 13. Configuration of the ground control points (GCPs) and check points in (**a**) Phantom2-Agriculture, (**b**) S1000-Agriculture-1, and (**c**) S1000-Agriculture-2 Datasets, and (**d**) a sample image of the targets established for both GCPs and check points in the test field.

Table 3. The number of utilized GCPs and check points, the derived σ_o, and the root-mean-square error (RMSE) values for Phantom2-Agriculture, S1000-Agiruculture-1, and S1000-Agriculture-2 Datasets.

	Phantom2-Agriculture	S1000-Agriculture-1	S1000-Agriculture-2
Number of GCPs	10	9	10
Number of Check Points	18	21	22
σ_o (pixel)	0.88	1.30	1.16
RMSE_X (m)	0.01	0.01	0.03
RMSE_Y (m)	0.01	0.01	0.03
RMSE_Z (m)	0.04	0.02	0.04
Extent of Covered Area	280 m × 100 m	650 m × 100 m	410 m × 100 m

4.3.2. Comparison with Airborne LiDAR Data

Since no ground control is available for **Phantom2-Building** Dataset, the accuracy evaluation is conducted through a comparison between the derived image-based sparse point cloud (See Figure 14a)

and the airborne LiDAR data, which was acquired by an Optech ALTM 3100 airborne laser scanning system. The airborne LiDAR data consists of multiple strips with an approximate 50% overlap. According to the manufacturer's specifications, horizontal accuracy of the acquired LiDAR data is less than 1/2000 of the flying height in meters while the vertical accuracy is less than 15 cm at a flying height of 1200 m and less than 25 cm at a flying height of 2000 m. A total number of 78,000 points has been cropped from the LiDAR data for the accuracy comparison. The average point spacing of the airborne LiDAR data is about 0.75 m (See Figure 14b). It is worth noting that although the absolute accuracy of airborne LiDAR is in the range of dozen centimeters, the relative accuracy within the test site covered by the building is much better (in the range of few centimeters). To perform the accuracy evaluation, the UAV image-based point cloud is precisely aligned with the LiDAR data through an ICPatch process [87]. Instead of using point point-to-point correspondence, the geometric primitives chosen in ICPatch are points and triangular patches. In this comparison, the UAV image-based point cloud is represented by the original points while the airborne LiDAR data is represented by the triangular patches from a TIN (Triangulated Irregular Network). The registration leads to a 5 cm RMSE value of the normal distance between the points and the corresponding triangular patches. Figure 14c illustrates the point-to-patch distances from the UAV image-based point cloud to the LiDAR data. Looking into Figure 14c, one can observe that there is a good compatibility between the two point clouds, especially in planar areas, such as building roof and flat ground.
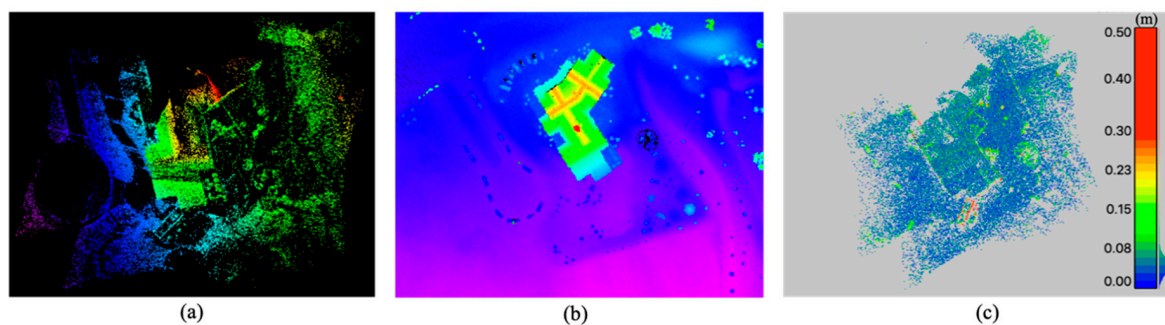

Figure 14. (**a**) The derived image-based sparse point cloud through the proposed automated aerial triangulation, (**b**) the utilized airborne LiDAR data, and (**c**) the point-to-patch distance after ICPatch registration.

4.4. Comparison with Pix4D

The last stage in the experimental tests is to compare the performance of the proposed approach with the Pix4D Mapper Pro software. In this research, both quantitative and qualitative evaluation methodologies have been established for the comparative performance analysis. It is worth noting that, to have a more general comparative analysis, we only compare our proposed framework to the default 3D Maps settings in Pix4D, which is mainly designed for vertical aerial images acquired using a grid flight plan with high overlap percentage. In addition, no geo-location information is used to facilitate the image matching process in Pix4D. The quantitative analysis is performed by comparing the number of images, whose EOPs have been successfully recovered in the proposed framework and the Pix4D software. Table 4 provides the number of estimated image EOPs for the four utilized experimental datasets, respectively. As can be seen in Table 4, the proposed automated aerial triangulation provides estimates for all the involved images in **Phantom2-Agriculture** Dataset, while only 487 out of 569 (86%) image EOPs are recovered in Pix4D. For the remaining datasets, both the proposed approach and Pix4D exhibit similar performance regarding the number of estimated image EOPs. Possible explanations for the inferior performance of **Phantom2-Agriculture** Dataset while using Pix4D software include the significant image distortions from the GoPro camera as well as the repetitive "texture" pattern within the agricultural field in question. Therefore, one can conclude that the proposed automated

aerial triangulation is more capable of dealing with UAV-based imagery with significant lens distortion and repetitive pattern.

Table 4. Number of images with estimated EOPs for Phantom2-Agriculture, Phantom2-Building, S1000-Agriculture-1, and S1000-Agriculture-2 datasets.

	Total Number of Images	**Number of Images with Estimated EOPs**	
		Proposed Approach	**Pix4D**
Phantom2-Agriculture	569	569	487
Phantom2-Building	81	81	81
S1000-Agriculture-1	421	418	420
S1000-Agriculture-2	639	639	639

The qualitative evaluation is conducted through the evaluation of the RGB-based orthophoto mosaic in the four utilized experimental datasets. In the proposed framework, a Digital Surface Model (DSM) is firstly interpolated from the UAV image-based point cloud. Then, the DSM together with the bundle adjustment-based EOPs as well as the estimated camera IOPs are used to generate the RGB-based orthophoto mosaic. For both the proposed framework and Pix4D, the spatial resolution for the DSM and the RGB-based orthophoto mosaic is set to 1 cm. Figures 15–18 illustrate the RGB-based orthophoto mosaic for the four experimental datasets, which are generated through the proposed framework and Pix4D, respectively. It is worth noting that since no color correction/balancing is utilized in the proposed framework, obvious boundaries among the set of mosaicked images can be observed in the derived orthophoto mosaic as shown in Figures 15–18. Compared to those Pix4D-based orthophoto mosaic with correctly balanced intensity values (see Figures 15–18), such boundaries are used to indicate the accuracy of the utilized image EOPs for the orthophoto generation. Specifically, when inaccurate image EOPs are used, obvious discrepancies can be observed along the boundaries among the mosaicked images. Carefully looking into these derived orthophotos, one can observe obvious gaps and discrepancies in the generated RGB-based orthophoto for **Phantom2-Agriculture** Dataset (see highlighted area in Figure 15b) while using Pix4D. For the same area, the proposed procedure, on the other hand, demonstrates better mosaicking quality in the final orthophoto as shown in Figure 15a. Such observation provides another evidence to support the claim that, compared to Pix4D, the proposed automated aerial triangulation has superior performance when dealing with images acquired with repetitive texture and significant radial lens distortions (e.g., **Phantom2-Agriculture** Dataset). For the remaining three experimental datasets, no significant differences can be observed in the RGB-based orthophoto that are respectively generated through the proposed procedure and the Pix4D software. In this regard, one can conclude that both the proposed approach and the Pix4D software are capable of providing comparable 3D reconstruction for UAV images, which are either acquired in a test site with "rich" texture (e.g., **Phantom2-Building**) or captured by a camera with less image distortions (e.g., **S1000-Agriculture-1** and **S1000-Agriculture-2** Datasets).
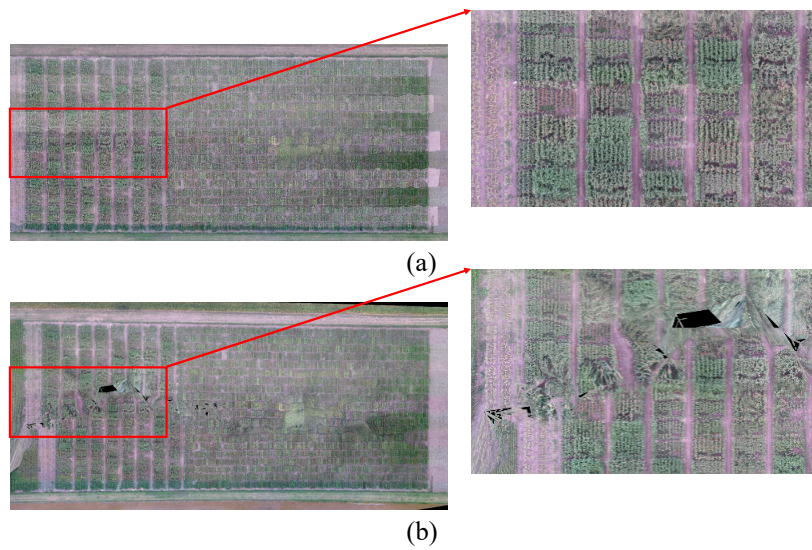
Figure 15. The generated RGB-based orthophoto mosaic and close-ups from (**a**) the proposed automated aerial triangulation, and (**b**) Pix4D for Phantom2-Agriculture Dataset.

Figure 16. The generated RGB-based orthophoto mosaic from (**a**) the proposed automated aerial triangulation, and (**b**) Pix4D for Phantom2-Building Dataset.
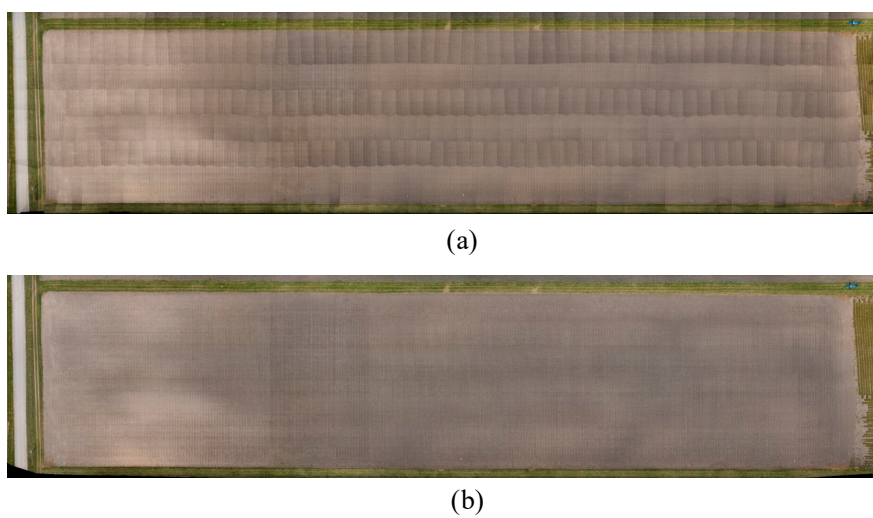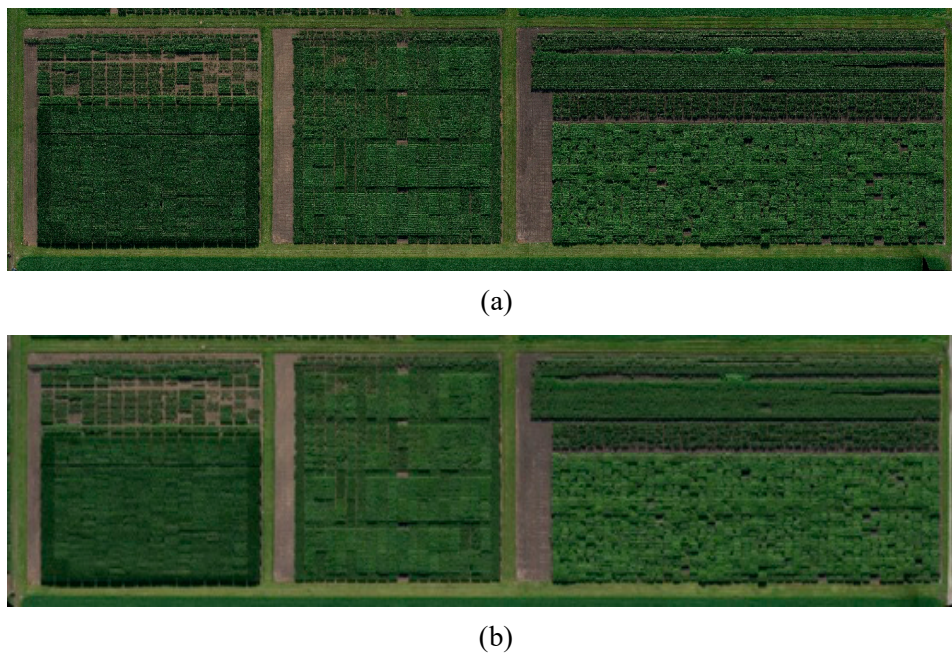

Figure 17. The generated RGB-based orthophoto mosaic from (**a**) the proposed automated aerial triangulation, and (**b**) Pix4D for S1000-Agriculture-1 Dataset.

(a)

(b)

Figure 18. The generated RGB-based orthophoto mosaic from (**a**) the proposed automated aerial triangulation, and (**b**) Pix4D for S1000-Agriculture-2 Dataset.

5. Conclusions and Recommendations for Future Work

This paper presents a fully automated framework for aerial triangulation of UAV-based images. Different from the existing commercial software, the proposed framework is a transparent system, and it can be incorporated with different user-defined constraints to improve the process of UAV image-based 3D reconstruction. In the proposed framework, two approaches which take advantage of prior information regarding the trajectory of the utilized UAV platform have been adopted for reliable ROP recovery of the involved stereo-images. Moreover, both incremental and global strategies have been proposed and investigated for the initial recovery of image EOPs. The performance of the proposed framework has been evaluated through four real image datasets acquired by two different UAV systems. The comparison between the incremental/global and the BA-based EOPs has shown the better accuracy and efficiency of the proposed global approach for the initial recovery of image EOPs. In terms of the accuracy of the derived 3D model, centimeter level accuracy (i.e., RMSE value < 5 cm) has been achieved in the proposed aerial triangulation framework for all the experimental datasets when compared to the GPS-surveyed check points/airborne LiDAR point cloud. In addition, both quantitative and qualitative evaluation has been established for the comparative analysis with Pix4D software. The evaluation has demonstrated the superior performance of the proposed framework when dealing with the acquired UAV images containing repetitive pattern and significant image distortions. Recommendation for future work includes focusing on the improvement of the proposed global approach for the initial recovery of image EOPs. It is worth noting that in the proposed global approach, the utilized multiple rotation averaging estimation ignores the orthogonality constraints during the estimated rotation matrices. Therefore, the comparison with other multiple rotation averaging techniques, such as the iterative approach introduced by Hartley, will be investigated for future work. In addition, an outlier detection/removal process, which aims at achieving more reliable parameter estimation, will be another focus for the global approach. Moreover, the augmentation of the proposed automated aerial triangulation framework with the available GNSS/INS information from the UAV platform will be also investigated. Finally, we will conduct the comparative analysis between our approach and the existing professional triangulation software, such as Trimble Inpho, for future research.

References

1. Candiago, S.; Remondino, F.; De Giglio, M.; Dubbini, M.; Gattelli, M. Evaluating multispectral images and vegetation indices for precision farming applications from UAV images. *Remote Sens.* **2015**, *7*, 4026–4047. [CrossRef]

2. Habib, A.; Han, Y.; Xiong, W.; He, F.; Zhang, Z.; Crawford, M. Automated ortho-rectification of UAV-based hyperspectral data over an agricultural field using frame RGB imagery. *Remote Sens.* **2016**, *8*, 796. [CrossRef]

3. Ribera, J.; He, F.; Chen, Y.; Habib, A.F.; Delp, E.J. Estimating phenotypic traits from UAV based RGB imagery. In Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Workshop on Data Science for Food, Energy, and Water, San Francisco, CA, USA, 13–17 August 2016.

4. Ribera, J.; Chen, Y.; Boomsma, C.; Delp, E.J. Counting Plants Using Deep Learning. In Proceedings of the 2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Montreal, QC, Canada, 14–16 November 2017.

5. 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-STARS* **2017**, *10*, 262–276. [CrossRef]

6. Chen, Y.; Ribera, J.; Boomsma, C.; Delp, E.J. Locating Crop Plant Centers from UAV-Based RGB Imagery. In Proceedings of the 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), Venice, Italy, 22–29 October 2017; pp. 2030–2037.

7. Chen, Y.; Ribera, J.; Boomsma, C.; Delp, E.J. Plant leaf segmentation for estimating phenotypic traits. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 3884–3888.

8. Habib, A.; Zhou, T.; Masjedi, A.; Zhang, Z.; Flatt, J.E.; Crawford, M. Boresight Calibration of GNSS/INS-Assisted Push-Broom Hyperspectral Scanners on UAV Platforms. *IEEE J-STARS* **2018**, *11*, 734–1749. [CrossRef]

9. Kim, D.-W.; Yun, H.S.; Jeong, S.-J.; Kwon, Y.-S.; Kim, S.-G.; Lee, W.S.; Kim, H.-J. Modeling and Testing of Growth Status for Chinese Cabbage and White Radish with UAV-Based RGB Imagery. *Remote Sens.* **2018**, *10*, 563. [CrossRef]

10. D'Oleire-Oltmanns, S.; Marzolff, I.; Peter, K.D.; Ries, J.B. Unmanned aerial vehicle (UAV) for monitoring soil erosion in Morocco. *Remote Sens.* **2012**, *4*, 3390–3416. [CrossRef]

11. Su, T.-C.; Chou, H.-T. Application of multispectral sensors carried on unmanned aerial vehicle (UAV) to trophic state mapping of small reservoirs: A case study of Tain-Pu reservoir in Kinmen, Taiwan. *Remote Sens.* **2015**, *7*, 10078–10097. [CrossRef]

12. Al-Rawabdeh, A.; He, F.; Moussa, A.; El-Sheimy, N.; Habib, A. Using an unmanned aerial vehicle-based digital imaging system to derive a 3D point cloud for landslide scarp recognition. *Remote Sens.* **2016**, *8*, 95. [CrossRef]

13. Fernández, T.; Pérez, J.L.; Cardenal, J.; Gómez, J.M.; Colomo, C.; Delgado, J. Analysis of landslide evolution affecting olive groves using UAV and photogrammetric techniques. *Remote Sens.* **2016**, *8*, 837. [CrossRef]

14. Hird, J.N.; Montaghi, A.; McDermid, G.J.; Kariyeva, J.; Moorman, B.J.; Nielsen, S.E.; McIntosh, A. Use of unmanned aerial vehicles for monitoring recovery of forest vegetation on petroleum well sites. *Remote Sens.* **2017**, *9*, 413. [CrossRef]

15. Tomaštík, J.; Mokroš, M.; Saloň, Š.; Chudý, F.; Tunák, D. Accuracy of photogrammetric UAV-based point clouds under conditions of partially-open forest canopy. *Forests* **2017**, *8*, 151. [CrossRef]

16. Fraser, B.T.; Congalton, R.G. Issues in Unmanned Aerial Systems (UAS) Data Collection of Complex Forest Environments. *Remote Sens.* **2018**, *10*, 908. [CrossRef]

17. George Pierce Jones, I.V.; Pearlstine, L.G.; Percival, H.F. An assessment of small unmanned aerial vehicles for wildlife research. *Wildl. Soc. Bull.* **2006**, *34*, 750–758. [CrossRef]

18. Hodgson, A.; Kelly, N.; Peel, D. Unmanned aerial vehicles (UAVs) for surveying marine fauna: A dugong case study. *PLoS ONE* **2013**, *8*, e79556. [CrossRef] [PubMed]

19. Fernández-Hernandez, J.; González-Aguilera, D.; Rodríguez-Gonzálvez, P.; Mancera-Taboada, J. Image-based modelling from unmanned aerial vehicle (UAV) photogrammetry: An effective, low-cost tool for archaeological applications. *Archaeometry* **2015**, *57*, 128–145. [CrossRef]

20. Jorayev, G.; Wehr, K.; Benito-Calvo, A.; Njau, J.; de la Torre, I. Imaging and photogrammetry models of Olduvai Gorge (Tanzania) by Unmanned Aerial Vehicles: A high-resolution digital database for research and conservation of Early Stone Age sites. *J. Archaeol. Sci.* **2016**, *75*, 40–56. [CrossRef]

21. He, F.; Habib, A.; Al-Rawabdeh, A. Planar constraints for an improved uav-image-based dense point cloud generation. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2015**, *40*, 269. [CrossRef]

22. Lari, Z.; Al-Rawabdeh, A.; He, F.; Habib, A.; El-Sheimya, N. Region-based 3D surface reconstruction using images acquired by low-cost unmanned aerial systems. *ISPRS-Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2015**, *40*, 167–173. [CrossRef]

23. 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]

24. Fraser, C.S. Digital camera self-calibration. *ISPRS J. Photogramm. Remote Sens.* **1997**, *52*, 149–159. [CrossRef]

25. Habib, A.; Morgan, M.F. Automatic calibration of low-cost digital cameras. *Opt. Eng.* **2003**, *42*, 948–956.

26. Cramer, M.; Stallmann, D.; Haala, N. Direct georeferencing using GPS/inertial exterior orientations for photogrammetric applications. *Int. Arch. Photogramm. Remote Sens.* **2000**, *33*, 198–205.

27. Skaloud, J. Direct georeferencing in aerial photogrammetric mapping. *Photogramm. Eng. Remote Sens.* **2002**, *68*, 207–209, 210.

28. Pfeifer, N.; Glira, P.; Briese, C. Direct georeferencing with on board navigation components of light weight UAV platforms. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2012**, *39*, 487–492. [CrossRef]

29. Remondino, F.; Barazzetti, L.; Nex, F.; Scaioni, M.; Sarazzi, D. UAV photogrammetry for mapping and 3d modeling–current status and future perspectives. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2011**, *38*, C22. [CrossRef]

30. Hartley, R.; Zisserman, A. *Multiple View Geometry in Computer Vision*; Cambridge University Press: Cambridge, UK, 2003.

31. Horn, B.K. Relative orientation. *Int. J. Comput. Vis.* **1990**, *4*, 59–78. [CrossRef]

32. Mikhail, E.M.; Bethel, J.S.; McGlone, J.C. *Introduction to Modern Photogrammetry*; Wiley: New York, NY, USA, 2001.

33. Longuet-Higgins, H.C. A computer algorithm for reconstructing a scene from two projections. *Nature* **1981**, *293*, 133. [CrossRef]

34. Hartley, R.I. In defense of the eight-point algorithm. *IEEE Trans. Pattern Anal. Mach. Intell.* **1997**, *19*, 580–593. [CrossRef]

35. Nistér, D. An efficient solution to the five-point relative pose problem. *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 756–770. [CrossRef]

36. Faugeras, O.D.; Maybank, S. Motion from point matches: Multiplicity of solutions. *Int. J. Comput. Vis.* **1990**, *4*, 225–246. [CrossRef]

37. Philip, J. A Non-Iterative Algorithm for Determining All Essential Matrices Corresponding to Five Point Pairs. *Photogramm. Rec.* **1996**, *15*, 589–599. [CrossRef]

38. Triggs, B. *Routines for Relative Pose of Two Calibrated Cameras from 5 Points*; Technical Report; INRIA, 2000.

39. Batra, D.; Nabbe, B.; Hebert, M. An alternative formulation for five point relative pose problem. In Proceedings of the Motion and Video Computing, IEEE Workshop on (WMVC), Austin, TX, USA, 23–24 February 2007; p. 21.

40. Kukelova, Z.; Bujnak, M.; Pajdla, T. Polynomial Eigenvalue Solutions to the 5-pt and 6-pt Relative Pose Problems. In Proceedings of the British Machine Vision Conference, Leeds, UK, 1–4 September 2008; pp. 56.1–56.10.

41. 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.

42. 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]

43. Ortin, D.; Montiel, J.M.M. Indoor robot motion based on monocular images. *Robotica* **2001**, *19*, 331–342. [CrossRef]

44. Fraundorfer, F.; Tanskanen, P.; Pollefeys, M. A minimal case solution to the calibrated relative pose problem for the case of two known orientation angles. In Proceedings of the 11th European conference on Computer vision, Crete, Greece, 5–11 September 2010; pp. 269–282.

45. Scaramuzza, D. Performance evaluation of 1-point-RANSAC visual odometry. *J. Field Robot.* **2011**, *28*, 792–811. [CrossRef]

46. Troiani, C.; Martinelli, A.; Laugier, C.; Scaramuzza, D. 2-point-based outlier rejection for camera-imu systems with applications to micro aerial vehicles. In Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, China, 31 May–7 June 2014; pp. 5530–5536.

47. Viéville, T.; Clergue, E.; Facao, P.D.S. Computation of ego-motion and structure from visual and inertial sensors using the vertical cue. In Proceedings of the 1993 (4th) International Conference on Computer Vision, Berlin, Germany, 11–14 May 1993; pp. 591–598.

48. Kalantari, M.; Hashemi, A.; Jung, F.; Guédon, J.-P. A new solution to the relative orientation problem using only 3 points and the vertical direction. *J. Math. Imaging Vis.* **2011**, *39*, 259–268. [CrossRef]

49. Naroditsky, O.; Zhou, X.S.; Gallier, J.; Roumeliotis, S.I.; Daniilidis, K. Two efficient solutions for visual odometry using directional correspondence. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 818–824. [CrossRef] [PubMed]

50. Robertson, D.P.; Cipolla, R. An Image-Based System for Urban Navigation. In Proceedings of the British Machine Vision Conference, London, UK, 7–9 September 2004; Volume 19, p. 165.

51. Gallagher, A.C. Using vanishing points to correct camera rotation in images. In Proceedings of the 2nd Canadian Conference on Computer and Robot Vision (CRV'05), Victoria, BC, Canada, 9–11 May 2005; pp. 460–467.

52. He, F.; Habib, A. Performance Evaluation of Alternative Relative Orientation Procedures for UAV-based Imagery with Prior Flight Trajectory Information. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *41*, 21–25. [CrossRef]

53. 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.

54. Fitzgibbon, A.W.; Zisserman, A. Automatic camera recovery for closed or open image sequences. In Proceedings of the 5th European Conference on Computer Vision, London, UK, 2–6 June 1998; pp. 311–326.

55. Hartley, R.I. Lines and points in three views and the trifocal tensor. *Int. J. Comput. Vis.* **1997**, *22*, 125–140. [CrossRef]

56. Agarwal, S.; Snavely, N.; Simon, I.; Seitz, S.M.; Szeliski, R. Building rome in a day. In Proceedings of the 2009 IEEE 12th International Conference on Computer Vision, Kyoto, Japan, 29 September–2 October 2009; pp. 72–79.

57. Frahm, J.-M.; Fite-Georgel, P.; Gallup, D.; Johnson, T.; Raguram, R.; Wu, C.; Jen, Y.-H.; Dunn, E.; Clipp, B.; Lazebnik, S. Building rome on a cloudless day. In Proceedings of the 11th European Conference on Computer Vision: Part IV, Crete, Greece, 5–11 September 2010; pp. 368–381.

58. Wu, C. Towards linear-time incremental structure from motion. In Proceedings of the 2013 International Conference on 3D Vision-3DV 2013, Seattle, WA, USA, 29 June–1 July 2013; pp. 127–134.

59. Schonberger, J.L.; Frahm, J.-M. Structure-from-motion revisited. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 4104–4113.

60. He, F.; Habib, A. Linear approach for initial recovery of the exterior orientation parameters of randomly captured images by low-cost mobile mapping systems. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2014**, *40*, 149. [CrossRef]

61. Hartley, R.; Trumpf, 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. Chatterjee, A.; Madhav Govindu, V. Efficient and robust large-scale rotation averaging. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013; pp. 521–528.

64. Carlone, L.; Tron, R.; Daniilidis, K.; Dellaert, F. Initialization techniques for 3D SLAM: A survey on rotation estimation and its use in pose graph optimization. In Proceedings of the 2015 IEEE International Conference on Robotics and Automation (ICRA), Seattle, WA, USA, 26–30 May 2015; pp. 4597–4604.

65. 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), Kauai, HI, USA, 8–14 December 2001; Volume 2.

66. Brand, M.; Antone, M.; Teller, S. Spectral solution of large-scale extrinsic camera calibration as a graph embedding problem. In Proceedings of the 8th European Conference on Computer Vision, Prague, Czech Republic, 11–14 May 2004; pp. 262–273.

67. Sinha, S.N.; Steedly, D.; Szeliski, R. A multi-stage linear approach to structure from motion. In Proceedings of the 11th European Conference on Trends and Topics in Computer Vision, Crete, Greece, 10–11 September 2010; pp. 267–281.

68. 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.

69. Cui, Z.; Jiang, N.; Tang, C.; Tan, P. Linear global translation estimation with feature tracks. In Proceedings of the British Machine Vision Conference (BMVC), Swansea, UK, 7–10 September 2015; pp. 41.6–46.13.

70. Cui, Z.; Tan, P. Global structure-from-motion by similarity averaging. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 864–872.

71. Jiang, N.; Cui, Z.; Tan, P. A global linear method for camera pose registration. In Proceedings of the 2013 IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013; pp. 481–488.

72. Förstner, W.; Wrobel, B.P. *Photogrammetric Computer Vision*; Springer: Berlin, Germany, 2016.

73. Granshaw, S.I. Bundle adjustment methods in engineering photogrammetry. *Photogramm. Rec.* **1980**, *10*, 181–207. [CrossRef]

74. Bartoli, A.; Sturm, P. Structure-from-motion using lines: Representation, triangulation, and bundle adjustment. *Comput. Vis. Image Understand.* **2005**, *100*, 416–441. [CrossRef]

75. Lee, W.H.; Yu, K. Bundle block adjustment with 3D natural cubic splines. *Sensors* **2009**, *9*, 9629–9665. [CrossRef] [PubMed]

76. Vo, M.; Narasimhan, S.G.; Sheikh, Y. Spatiotemporal bundle adjustment for dynamic 3d reconstruction. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1710–1718.

77. Triggs, B.; McLauchlan, P.F.; Hartley, R.I.; Fitzgibbon, A.W. Bundle adjustment—A modern synthesis. In *International Workshop on Vision Algorithms*; Springer: Berlin, Germany, 1999; pp. 298–372.

78. Lourakis, M.I.; Argyros, A.A. SBA: A software package for generic sparse bundle adjustment. *ACM Trans. Math. Softw. (TOMS)* **2009**, *36*, 2. [CrossRef]

79. Wu, C.; Agarwal, S.; Curless, B.; Seitz, S.M. Multicore bundle adjustment. In Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence, RI, USA, 20–25 June 2011; pp. 3057–3064.

80. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]

81. He, F.; Habib, A. Automatic orientation estimation of multiple images with respect to laser data. In Proceedings of the ASPRS 2014 Annual Conference, Louisville, KY, USA, 23–24 Match 2014.

82. Horn, B.K. Closed-form solution of absolute orientation using unit quaternions. *JOSA A* **1987**, *4*, 629–642. [CrossRef]

83. Guan, Y.; Zhang, H. Initial registration for point clouds based on linear features. In Proceedings of the 2011 Fourth International Symposium on Knowledge Acquisition and Modeling, Sanya, China, 8–9 October 2011; pp. 474–477.

84. He, F.; Habib, A. A closed-form solution for coarse registration of point clouds using linear features. *J. Surv. Eng.* **2016**, *142*, 04016006.

85. Watson, G.A. Computing helmert transformations. *J. Comput. Appl. Math.* **2006**, *197*, 387–394. [CrossRef]

86. 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.

87. Habib, A.; Detchev, I.; Bang, K. A comparative analysis of two approaches for multiple-surface registration of irregular point clouds. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2010**, *38*, 61–66.