Automated Relative Orientation of UAV-Based Imagery in the Presence of Prior Information for the Flight Trajectory

Fangning He and Ayman Habib

Abstract

UAV-based 3D reconstruction has been used in various applications. However, mitigating the impact of outliers in automatically matched points remains to be a challenging task. Assuming the availability of prior information regarding the UAV trajectory, this paper presents two approaches for reliable estimation of Relative Orientation Parameters (ROPs) in the presence of high percentage of matching outliers. The first approach, which assumes that the UAV platform is moving at a constant flying height while maintaining the camera in a nadir-looking orientation, provides a two-point closed-form solution. The second approach starts from prior information regarding the flight trajectory to define a linearized model, which is augmented with a built-in outlier removal procedure, to estimate a refined set of ROPs. Experimental results from real datasets demonstrate the feasibility of the proposed approaches in providing reliable ROPs from UAV-based imagery in the presence of a high percentage of matching outliers (up to 90 percent).

Introduction

Accurate 3D modeling has become a key prerequisite for several applications, such as urban planning, archaeological documentation, environmental monitoring, disaster aftermath assessment, change detection, precision agriculture, and military applications. 3D reconstruction/representation of objects can be achieved through either active or passive remote sensing systems. Due to financial and technical constraints, passive sensor systems, which commonly use digital line/frame imaging sensors, are still an optimum option for various 3D reconstruction applications (Remondino and El-Hakim, 2006). Within the photogrammetric research community, automation of image-based 3D reconstruction has been investigated for decades. In order to derive high-quality 3D reconstruction, conventional photogrammetric mapping requires the knowledge of the Interior Orientation Parameters (IOPs) of the utilized cameras, Exterior Orientation Parameters (EOPs) of the involved images, and corresponding points/features within overlapping images. The IOPs can be derived from a camera calibration process (Fraser, 1997; Habib and Morgan, 2003). The EOPs of the involved imagery can be either derived through an indirect or a direct geo-referencing process (Cramer *et al.*, 2000; Skaloud, 2002). For indirect geo-referencing, the image EOPs are indirectly established using tie and control points. However, the identification of reliable tie points and the set-up of control points are time-consuming and costly activities. In spite of the fact that the direct geo-referencing simplifies the derivation of the EOPs for each exposure station, significant initial investment for the acquisition of a high-end GNSS/INS Position and Orientation System (POS) is required, especially when seeking high level of reconstruction accuracy. For 3D reconstruction, whether it is based on indirect or direct georeferencing, we need to automatically identify conjugate points in overlapping images (this is commonly known as the matching problem). Image matching can be a challenging task when dealing with imagery that has poor and/or repetitive texture. Therefore, one can argue that the adoption of conventional photogrammetric mapping techniques for image-based 3D reconstruction, especially for some emerging applications such as precision agriculture, can be limited.

Large-area 3D reconstruction has been traditionally established using manned-airborne data acquisition platforms. Unmanned Aerial Vehicles (UAVs) have recently emerged as a promising geospatial data acquisition system. This promise is mainly attributed to recent advances in low-cost direct georeferencing systems as well as imaging sensors operating at different portions of the electromagnetic spectrum. Compared to manned-airborne systems, the advantages of UAVs include their low-cost, ease of storage and deployment, ability to fly lower and collect high resolution data with consumer-grade cameras, and filling an important gap between wheel-based and manned-airborne platforms. To date, several research efforts have been geared towards the use of UAVs for small-area mapping applications (He *et al.*, 2015; He and Habib, 2014; Lari *et al.*, 2015). Structure from Motion (SfM), which was initiated by the computer vision research community, has been widely adopted for UAV-based 3D reconstruction. Similar to the procedure that has been adopted by the photogrammetric community for decades, SfM is implemented in three steps to simultaneously estimate the EOPs of the involved images and drive 3D coordinates of matched features within the overlap area (Hartley and Zisserman, 2003; Huang and Netravali, 1994). In the first step, the relative orientation parameters (ROPs) relating stereo-images are initially estimated using automatically identified conjugate point and/or line features. 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. The bundle adjustment procedure can achieve the best 3D reconstruction accuracy provided that we are utilizing sufficiently accurate approximations of the unknowns and correct feature correspondences (Stewenius *et al.*, 2006). In this regard, one should note that accurate estimation of the ROPs, which define the position and orientation of one image relative to another, is a prerequisite for any 3D reconstruction using SfM (Horn, 1990a).

In the past few decades, recovery of the ROPs has been investigated within the photogrammetric and computer vision

Lyles School of Civil Engineering, Purdue University, West Lafayette, IN 47906 (he270@purdue.edu).

Photogrammetric Engineering & Remote Sensing
Vol. 82, No. 11, November 2016, pp. 879–891.
0099-1112/16/879–891
© 2016 American Society for Photogrammetry and Remote Sensing
doi: 10.14358/PERS.82.11.879

research communities (Habib and Kelley, 2001; Heipke, 1997; Tang and Heipke, 1996; Zhang *et al.*, 2011). In general scenarios, the IOPs of the utilized camera are usually assumed to be known. 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 (Mikhail *et al.*, 2001), where a least-squares adjustment is solved while using a minimum of five conjugate points. However, due to the nonlinear nature of the co-planarity model, approximate values for the unknowns, which are refined through an iterative process, have to be available. Establishing good-quality approximations can be challenging for situations where the mapping platform exhibits excessive maneuvers between the data acquisition epochs (e.g., close range mapping applications). To address such a challenge, several closed-form solutions, which do not require approximations, for ROP recovery have been developed, such as the eight-point and the five-point algorithms (Hartley, 1997; Longuet-Higgins, 1987; Nistér, 2004).

The abovementioned ROP recovery procedures are based on reliable conjugate point pairs 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 necessary to augment the relative orientation procedure with strategies for outlier removal. RANSAC (Random Sample Consensus) is a commonly-used strategy to filter out outliers during a model-fitting procedure, in general, and ROP recovery, in particular (Fischler and Bolles, 1981). RANSAC works by conducting random draws of the necessary samples for ROP estimation, i.e., five or eight hypothesized conjugate pairs for the five-point and eight-point algorithms, respectively, and identifying the compatible matches with the estimated parameters from the different draws. The draw that results in the largest consensus is used together with the compatible matches to derive a reliable estimate of the ROPs. In spite of 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 recovery when there is a false hypothesis providing larger consensus. In addition to the traditional RANSAC approach for the five-point or eight-point algorithm, ROP recovery while considering prior information regarding the flight trajectory has been utilized to reduce the number of required conjugate pairs (Faugeras and Maybank, 1990; Ortin and Montiel, 2001; Scaramuzza, 2011a; Scaramuzza *et al.*, 2009). Such approaches, which were mainly initiated by the mobile robotics research community, assume the availability of some constraints on the system trajectory during data acquisition. ROP recovery while using fewer samples would require fewer trials to identify the largest consensus (Troiani *et al.*, 2014).

ROP recovery in the presence of prior information regarding the system trajectory has been mainly focused on indoor and outdoor terrestrial mobile mapping systems. However, dedicated approaches for UAV-based mobile mapping while considering the possibility of having high percentage of outliers are still missing. In this paper, we are introducing two approaches for the recovery of ROPs between stereo images captured by UAV-platforms in the presence of prior information regarding the system trajectory. These approaches have the following characteristics:

- The first approach assumes that the involved images are acquired by a UAV platform moving at constant flying height while operating a nadir-looking camera. Such constraints are considered to establish a two-point closed-form algorithm for ROP recovery. This approach is integrated within a RANSAC framework for outlier removal.
- The second approach assumes the availability of approximate values for the ROPs, which can be acquired from onboard consumer-grade navigation sensors. Then, The ROPs are iteratively estimated through a linearized co-planarity model with a built-in outlier detection/removal process.
- Since both approaches are coupled with a built-in outlier detection/removal process, they can deal with ROP recovery, where the initial point correspondences are contaminated with a high percentage of outliers.

This paper starts with a literature review of related work, which is followed by the mathematical details of the two proposed approaches for ROP recovery in the presence of prior information for the system trajectory. Afterwards, experimental results using real datasets are presented. Finally, drawn conclusions and recommendations for future work are introduced.

Related Work

As mentioned earlier, several closed-form solutions, which do not require approximations for ROP recovery, have been developed. Motivated by the concept of the Essential matrix, which encapsulates the epipolar geometry relating stereo-images, an eight-point algorithm was proposed by Longuet-Higgins (1987) for recovering the structure of a scene from two views that have been captured by a calibrated camera. In spite of its simplicity, such eight-point algorithm does not consider the constraints among the nine elements of the Essential matrix (i.e., constraints should be imposed to consider the fact that those elements are defined by five independent parameters). Thus, it is 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 the eight-point algorithm was proposed by Hartley (1997), 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 non-linear algorithm. Given that a minimum of five conjugate point pairs are needed for ROP recovery, several five-point algorithms have been proposed as alternatives to the eight-point algorithm (Faugeras and Maybank, 1990; Philip, 1996; Triggs, 2000). The most efficient five-point algorithm is the one proposed by Nistér (2004), and later improved by Stewenius *et al.* (2006). Compared to the original eight-point algorithms, the great advantage of five-point algorithms is that they take into account the epipolar geometry to enforce constraints on the elements of the Essential matrix. Therefore, five-point algorithms can better handle noisy image measurements and planar scenes (Philip, 1998).

Meanwhile, in recent years, motivated by the availability of consumer-grade navigation systems, several research efforts have been exerted towards ROP recovery while taking advantage of prior information regarding the system trajectory during data acquisition. In the existing body of literature, one or more of the ROPs between the images of a stereo-pair are assumed to be known. Troiani *et al.* (2014) introduced a two-point algorithm 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. In this work, the translation parameters are linearly recovered using two point correspondences. Traditionally, three rotation

angles are needed to define the rotation matrix relating the images of a stereo-pair. Since a rotation matrix can be alternatively defined by a rotation angle around an axis in space, the relative rotation between stereo-images sharing the same axis of rotation, which is usually denoted reference direction, can be defined by a single parameter (Viéville *et al.*, 1993). For such situations, several three-plus-one algorithms, which utilize three point correspondences and a known reference direction, have been developed as a substitute of the classical five-point algorithm (Kalantari *et al.*, 2011; Naroditsky *et al.*, 2012; Robertson and Cipolla, 2004). Prior information regarding the reference direction can be either derived from a detected vanishing point (Gallagher, 2005; Hartley and Zisserman, 2003) or using a gravity sensor onboard mobile mapping platforms, where the gravity vector becomes the reference direction. A recent three-point solution has been proposed by Fraundorfer *et al.* (2010). In their work, a simplified Essential matrix is estimated from three point correspondences using two known rotation angles, which are acquired from a Smartphone. Experimental results demonstrate that the proposed three-point algorithm can efficiently cope with planar scenes and even three collinear image points.

Additional geometric constraints regarding the system trajectory have been also considered for ROP recovery. For example, Ortin and Montiel (2001) assume that the platform's movement is constrained to a horizontal plane, and the rotation of the utilized camera is constrained to an axis orthogonal to the plane of motion (i.e., the normal to the plane of motion defines the reference direction). Using such assumptions, they conclude that two corresponding points are enough for determining the ROPs relating a stereo-pair. Scaramuzza (2009 and 2011b) demonstrates that the movement of wheel-based vehicles can be considered as locally circular and planar. As a result, a one-point solution with a single iteration outlier detection can be adopted. A comparison of the five, two, one-point algorithms incorporated within RANSAC for outlier removal has been conducted by Scaramuzza (2011a). Since the manipulation of stereo-images captured by UAVs in the presence of high percentage of outliers while considering prior information regarding the system trajectory has not been addressed, this paper is dedicated to addressing such scenario by presenting two alternative approaches.

Methodology for the Proposed Approaches

Current UAV-based mapping is usually executed according a mission plan while relying on a consumer-grade navigation sensor within the platform's autopilot. Therefore, prior information, which describes the trajectory of the platform including the involved cameras, can be utilized to facilitate ROP recovery for stereo-images. In this research work, we assume that such prior information can be either derived from a specific flight configuration (e.g., a UAV platform moving at a constant flying height while operating a nadir-looking camera) or established by a low-cost Micro-Electro-Mechanical-System (MEMS) integrated with a GPS unit onboard the UAV. In this section, the proposed approaches that take advantage of such prior information are presented. First, we introduce the conceptual basis for ROP recovery while using the Essential matrix. Then, a two-point approach integrated within RANSAC for outlier removal and reliable ROP estimation is presented. This approach is based on a specific flight configuration, where two rotation angles and one of the translation parameters are assumed to be zero. Then, a second approach, which is based on a linearized co-planarity model, is introduced. In this approach, the ROPs are iteratively updated through a built-in outlier detection strategy until a pre-defined stopping criterion is satisfied. Both approaches can be used to determine an estimate for the ROPs using a set of initial matches, which contain outliers. These matches could be based on a SIFT detector and descriptor that has been applied to stereo-images and using the Euclidean distances between the SIFT descriptors for the detected features in both images to identify potential matches (Lowe, 2004).

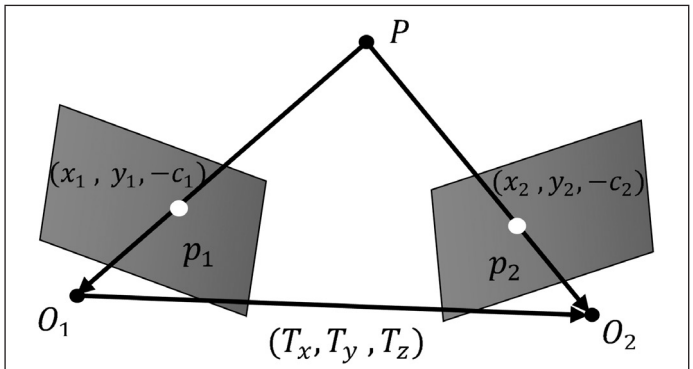

Figure 1. The co-planarity model relating stereo-images.

Conceptual Basis for the Essential Matrix

The Essential matrix E describes the epipolar geometry relating two corresponding points in a stereo-pair (Longuet-Higgins, 1987). The Essential matrix is based on the co-planarity constraint, which is depicted in Figure 1. The co-planarity constraint mathematically describes the fact that an object point P, the corresponding image points, and the two perspective centers O_1 and O_2 of a stereo-pair must lie on the same plane (Equation 1).

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

In this equation, p_1 and p_1 are two corresponding points, where $p = (x, y, -c)^T$ represents the image coordinates corrected for the principal point offset and camera-specific distortions. In this research work, we consider the case that the stereo-images are acquired by two different cameras. The rotation matrix R, which is defined by three rotation angles ω, ϕ, and κ, describes the relative rotation relating overlapping images. $\vec{T}$ is the translation vector describing the baseline between the stereo-images, and it can be defined by three translation components (T_x, T_x, T_z). The cross product in Equation 1 can be simplified using the skew-symmetric matrix $\hat{T}$ in Equation 2. More specifically, the 3-by-3 matrix $\hat{T}$ simplifies the cross product of two vectors to a matrix-vector multiplication. Using Equation 2, one can derive the expression for the Essential matrix as shown in Equation 3:

$$p_1^{\mathrm{T}} \hat{T} R p_2 = 0 \tag{2}$$

where,

$$\hat{T} = \begin{bmatrix} 0 & -Tz & Ty \\ Tz & 0 & -Tx \\ -Ty & Tx & 0 \end{bmatrix}$$

$$E = \begin{bmatrix} 0 & -Tz & Ty \\ Tz & 0 & -Tx \\ -Ty & Tx & 0 \end{bmatrix} R = \begin{bmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & e_{33} \end{bmatrix} \tag{3}$$

The nine elements of the Essential matrix are defined by the five elements of the ROPs (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. Such constraints are explained as follows:

1. The Essential matrix has rank two. Therefore, its determinant has to be zero as shown in Equation 4, which leads to what is known as the cubic constraint on the nine unknown parameters of the Essential matrix.

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

2. The Essential matrix has two equal non-zero singular values, which leads to the trace constraint as represented by Equation 5. Given the rank of , two independent equations on the nine unknown parameters can be deduced from the equality in Equation 5 (Horn, 1990b).

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

3. The nine elements of the Essential matrix can be only determined up to a scale, which provides the fourth constraint.

Considering these four constraints, one can conclude that a minimum of five-point-correspondence is sufficient to estimate the nine unknown parameters of the Essential matrix. Using five point correspondences, a system of five linear equations of the form in Equation 6 can be established. An efficient solution of the Essential matrix using five feature correspondences while considering the above constraints is provided by Nistér (2004). The derived Essential matrix can then be used to derive the rotation matrix R and the translation vector $\overline{T}$ relating the stereo-images. Possible solutions for R and $\overline{T}$ from a given Essential matrix have been introduced by Horn (1990b) and Nistér (2004):

$$\begin{aligned} &x_1x_2e_{11} + x_1y_2e_{12} - x_1c_2e_{13} + y_1x_2e_{21} + y_1y_2e_{22} - y_1c_2e_{23} \\ &-c_1x_2e_{31} - c_1y_2e_{32} + c_1c_2e_{33} = 0 \end{aligned} \tag{6}$$

Two-Point Approach for Relative Orientation Recovery

This section introduces the proposed two-point approach for ROP recovery while considering prior information regarding the platform trajectory. First, the mathematical model that considers the geometric constraints arising from the assumed motion trajectory on the Essential matrix is introduced. Then, a closed-form solution, which utilizes two point correspondences, is presented. Finally, the integration of the two-point algorithm within RANSAC for outlier removal is outlined.

Derivation of the Two-Point Approach

This approach is based on acquired imagery from a nadir-looking camera onboard a UAV platform moving at a constant flying height. Within the robotics research community, this flight configuration is known as "planar motion," where the platform's motion is constrained to a horizontal plane, and the rotation of the image plane is constrained along an axis orthogonal to the horizontal plane (i.e., with a reference direction that coincides with the normal to the plane of motion). The UAV-based planar motion leads to two geometric constraints that can be used to reduce and simplify the elements of the Essential matrix. As can be seen in Figure 2, the geometric constraints lead to the following:

1. For a nadir-looking camera, the rotation angles ω and ϕ are assumed to be zero.

   Therefore, the relative rotation of the camera between the images of a stereo-pair is constrained to the rotation angle κ (i.e., heading);

2. For a planar motion along the horizontal plane, the T_Z translation component is assumed to be zero.

Therefore, for a planar motion, the rotation matrix R and translation vector $\overline{T}$ relating the stereo-images can be expressed according to Equation 7, where κ is the rotation angle, and T_x and T_y are the translation components describing the horizontal planar motion of the UAV platform. The expressions for R and $\overline{T}$ can be substituted into Equation 3. This substitution will lead to the simplified Essential matrix in Equation 8, where L_1, L_2, L_3, and L_4 are used to denote the four unknown parameters of the Essential matrix E. As can be seen in Equation 8, 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 constraint relating the four elements of the Essential matrix. A closer inspection of the relationships between (L_1, L_2, L_3, L_4) and (T_x, T_y, and κ), one can introduce the constraint in Equation 9.

$$R = \begin{bmatrix} \cos\kappa & -\sin\kappa & 0 \\ \sin\kappa & \cos\kappa & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } T = \begin{bmatrix} T_x \\ T_y \\ 0 \end{bmatrix} \tag{7}$$

$$\begin{aligned} E &= \begin{bmatrix} 0 & 0 & Ty \\ 0 & 0 & -Tx \\ -Ty & Tx & 0 \end{bmatrix} \begin{bmatrix} \cos\kappa & -\sin\kappa & 0 \\ \sin\kappa & \cos\kappa & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 & T_y \\ 0 & 0 & -T_x \\ -T_y\cos\kappa + T_x\sin\kappa & T_y\sin\kappa + T_x\cos\kappa & 0 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 & L_1 \\ 0 & 0 & L_2 \\ L_3 & L_4 & 0 \end{bmatrix} \end{aligned} \tag{8}$$

$$L_1^2 + L_2^2 - L_3^2 - L_4^2 = 0. \tag{9}$$

Given the simplified form for the Essential matrix that describes the relationship between conjugate points in stereo-images captured under horizontal planar motion constraints, we will focus on establishing the closed-form solution of the proposed two-point approach. Using the simplified Essential matrix, one can expand the relationship between the image coordinates of conjugate points to the form in Equation 10, where the image coordinates of the conjugate points p_1 and p_2 are represented by (x_1, y_1, $-c_1$) and (x_2, y_2, $-c_2$) after correcting for the principal point offsets and camera-specific distortions.

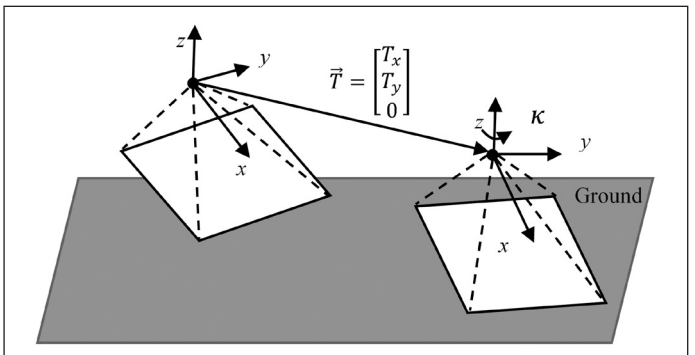

Figure 2. Stereo-images from a UAV platform equipped with a nadir-looking camera while moving at a constant flying height.

$$p_1^{\mathrm{T}}\begin{bmatrix}0 & 0 & L_1\\ 0 & 0 & L_2\\ L_3 & L_4 & 0\end{bmatrix}p_2 = -x_1c_2L_1 - y_1c_2L_2 - x_2c_1L_3 - y_2c_1L_4 = \mathbf{0} \quad (10)$$

At this stage, the ROP recovery should be concerned with coming up with an estimate of the four parameters (L_1, L_2, L_3, L_4) while observing the inherent constraint among these parameters, which is represented by Equation 9, and the fact that these parameters can be only determined up to an arbitrary scale. Therefore, two conjugate point pairs should be sufficient for deriving the simplified Essential matrix. The proposed closed-form solution for deriving the elements of the Essential matrix estimation starting from Equation 10 can be carried out as follows:

1. Given two conjugate point pairs, two linear equations, which can be represented using the matrix form in Equation 11, can be established.

$$\begin{bmatrix}-x_1^1c_2 & -y_1^1c_2 & -x_2^1c_1 & -y_2^1c_1\\ -x_1^2c_2 & -y_1^2c_2 & -x_2^2c_1 & -y_2^2c_1\end{bmatrix}\begin{bmatrix}L_1\\ L_2\\ L_3\\ L_4\end{bmatrix} = A_{2\times4}L_{4\times1} = \mathbf{0} \quad (11)$$

   where,

   $\left(x_1^1, y_1^1, -c_1\right)$ and $\left(x_2^1, y_2^1, -c_2\right)$ are the image coordinates for the first conjugate point pair, and $\left(x_1^2, y_1^2, -c_1\right)$ and $\left(x_2^2, y_2^2, -c_2\right)$ are the image coordinates for the second conjugate point pair.
2. As can be seen in Equation 11, the $L_{4\times1}$ vector belongs to the null space of the $A_{2\times4}$ matrix, which has a rank of two and is comprised of the image coordinates of available conjugate point pairs. Therefore, the $L_{4\times1}$ vector can be derived through a linear combination of the basis vectors $\tilde{X}$, and $\tilde{Y}$ spanning the right null-space of the matrix $A_{2\times4}$.

   Assuming that the basis vectors $\tilde{X}$, and $\tilde{Y}$ are presented by the elements in Equation 12, L_1, L_2, L_3, L_4 can be derived through the linear combination in Equation 13, where α and b are two arbitrary scale factors. Since the Essential matrix is determined only up to an arbitrary scale, one can choose a value of 1 for the scale factor *b*. Thus, the Essential matrix can be expressed by the form in Equation 14 using the basis vectors spanning the right null space of the $A_{2\times4}$ matrix.

Using the expressions for L_1, L_2, L_3, L_4 in Equation 13 and the inherent constraint in Equation 9, one can derive the second order polynomial in the unknown scale factor in Equation 15. The second order polynomial in Equation 15 provides up to two estimates for the scale factor α. Thus, two Essential matrices might be derived. For a given Essential matrix, R and T can be recovered through either the introduced singular value decomposition (SVD) approach by Nistér (2004) or the proposed closed-form solution by Horn (1990b). As presented by Horn (1990b), a total of four possible solutions of the rotation matrix R and translation vector $\overline{T}$ can be recovered from a single Essential matrix. Therefore, up to eight solutions for R and $\overline{T}$ can be derived from this approach. In order to identify the valid Essential matrix among the available solutions, two additional constraints can be utilized as follows:

- The light rays connecting a derived object point and perspective centers should be on the same side of the baseline.
- The derived object points should be below the camera stations.

$$\tilde{X}=[\tilde{X}_1, \tilde{X}_2, \tilde{X}_3, \tilde{X}_4]^4 \text{ and } \tilde{Y}=[\tilde{Y}_1, \tilde{Y}_2, \tilde{Y}_3, \tilde{Y}_4]^{\mathrm{T}} \quad (12)$$

$$\begin{bmatrix}L_1\\ L_2\\ L_3\\ L_4\end{bmatrix} = a\begin{bmatrix}\tilde{X}_1\\ \tilde{X}_2\\ \tilde{X}_3\\ \tilde{X}_4\end{bmatrix} + b\begin{bmatrix}\tilde{Y}_1\\ \tilde{Y}_2\\ \tilde{Y}_3\\ \tilde{Y}_4\end{bmatrix} \quad (13)$$

$$E = \begin{bmatrix}0 & 0 & L_1\\ 0 & 0 & L_2\\ L_3 & L_4 & 0\end{bmatrix} = \begin{bmatrix}0 & 0 & a\tilde{X}_1 + \tilde{Y}_1\\ 0 & 0 & a\tilde{X}_2 + \tilde{Y}_2\\ a\tilde{X}_3 + \tilde{Y}_3 & a\tilde{X}_4 + \tilde{Y}_4 & 0\end{bmatrix} \quad (14)$$

$$\left(\tilde{X}_1^{\,2} + \tilde{X}_2^{\,2} - \tilde{X}_3^{\,2} - \tilde{X}_4^{\,2}\right)a^2 + 2\left(-\tilde{X}_1\tilde{Y}_1 - \tilde{X}_2\tilde{Y}_2 + \tilde{X}_3\tilde{Y}_3 + \tilde{X}_4\tilde{Y}_4\right)a + (\tilde{Y}_1^{\,2} + \tilde{Y}_2^{\,2} - \tilde{Y}_3^{\,2} - \tilde{Y}_4^{\,2}) = 0 \quad (15)$$

In summary, the proposed two-point approach assumes that the involved images are acquired from a nadir-looking camera onboard a UAV platform moving at a constant flying height. Therefore, the ω and ϕ rotation angles and the T_Z translation component are assumed to be zero. Such prior flight information leads to the fact that a minimum of two conjugate point pairs can be used to derive the Essential matrix relating a stereo-pair through a closed form. In contrast to the two-point algorithm proposed by (Ortin and Montiel, 2001), where a non-linear optimization is adopted, the two-point approach introduced in this paper provides a direct solution for the Essential matrix without the need for any iterative optimization. Moreover, the proposed two-point approach can be used while having more than two conjugate point pairs. In this case, the matrix A as in Equation 11 is first generated using all conjugate point pairs. Then, the vectors $\tilde{X}$, and $\tilde{Y}$ corresponding to the two smallest singular values of the matrix A or the product A^TA are evaluated and used to derive the Essential matrix. One should note that the proposed two-point approach can deal with stereo pairs that have any heading angles and planimetric translation components. However, it cannot tolerate significant variations in the tilt angles as well as flying height between the stereo images.

RANSAC *Implementation for Outlier Removal*

Similar to previous five-point/eight-point algorithms for ROP recovery, the proposed two-point approach can be integrated within a RANSAC framework for outlier detection and removal. More specifically, a random sample comprised of two conjugate point pairs is first drawn from potential matches and used to derive the Essential matrix E according to the procedure proposed in the previous section. To derive other matches that are compatible with such estimate of the Essential matrix (i.e., derive the corresponding inliers), the Sampson distance (Hartley and Zisserman, 2003), which is the first order approximation of the normal distances between a given conjugate point pair and the respective corresponding epipolar lines, is evaluated for the remaining potential matches. Such a sampling-and-testing procedure is repeated until a required number of trials/draws is achieved. The required number of trials, which is based on an assumed percentage of inliers, is derived according to a probabilistic basis to ensure that at least one correct draw has been executed as seen in Equation 16. Finally, the random sample with the highest number of inliers is selected and used for ROP estimation. In

order to evaluate the computational efficiency of the proposed two-point approach when coupled with RANSAC for outlier removal in contrast to three, five, and eight point algorithms, the required number of trials N is presented in Table 1 (where ε is assumed to be 0.5 and the probability of having at least a single two correct conjugate point pairs within these trials is set to 0.99). As can be seen in Table 1, the required number of RANSAC trials is significantly reduced by using fewer number of conjugate point pairs. Therefore, the proposed two-point approach with a built-in RANSAC outlier detection/removal process is advantageous when compared with other approaches that require larger number of conjugate point pairs.

$$N = \frac{\log(1-p)}{\log(1-(1-\varepsilon)^s)} \tag{16}$$

where s is the minimum number of required conjugate point pairs for ROP estimation; ε is the probability of choosing an incorrect conjugate point pair, which is equivalent to the ratio between the assumed number of incorrect conjugate point pairs and the total number of potential conjugate point pairs; $(1-\varepsilon)^s$ is the probability of having correct conjugate point pairs in a single draw, and p is the probability of having at least a single sample comprised of correct conjugate point pairs.

TABLE 1. NUMBER OF RANSAC TRIALS FOR DIFFERENT SAMPLE SIZES OF CONJUGATE POINT PAIRS

Number of required conjugate point pairs (s)	2	3	5	8
Number of Trials (N)	16	35	145	1,177

Iterative Approach for Relative Orientation Recovery

This 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. The platform trajectory could be available from a designed mission plan and/or a consumer-grade GNSS/INS unit onboard the UAV. The co-planarity model is simplified to a linear equation involving the corrections to the trajectory-based approximate values for the ROPs as well as the image coordinates of conjugate point pairs. The corrections are then used to refine the approximate ROPs. This process is iteratively executed while removing potential outliers through the iteration procedure until a convergence criterions is achieved.

Mathematical Model

Given approximate values for the platform's rotation matrix R and translation $\bar{T}$ between the images of a stereo-pair and assuming unknown incremental rotation and translation corrections (δR *and* δT), the co-planarity model can be represented by Equation 17, which can be expanded to the form in Equation 18. Assuming that we have small deviations from the assumed flight trajectory, the incremental rotation matrix δR can be represented by Equation 19. The individual terms in Equation 18 can be combined into the terms M_1, M_2, and M_3 as seen in Equation 20. Since for the relative orientation, we can only estimate the translation vector up to an arbitrary scale, the correction to the translation component δT_x can be set to zero. In this regard, one should note that setting to zero implicitly assumes that the system trajectory is mainly aligned along the x-axis of the image coordinate system. Such assumption will limit the flight configurations that we can handle through this approach. The mitigation of such strict assumption will be discussed later in this section. Using the assumptions of small deviations from the given trajectory as well as having a flight trajectory that is mainly aligned along the x-axis, Equation 20 can be expanded to the form in Equation 21 after ignoring 2nd order correction terms (e.g., $\Delta\omega\delta T_y$ are assumed to be very small and set to zero). In Equation 21, $\Delta\omega$, $\Delta\phi$, $\Delta\kappa$, δT_y, and δT_z are the unknown incremental corrections to the approximate ROPs. Given five or more conjugate point pairs, one can derive a Least Squares Adjustment (LSA) – based closed form for the estimation of the unknown corrections, which can be used to derive refined estimates for the relative rotation and translation between the images of a stereo pair as seen in Equation 22. The refined parameters can be then used as a better estimate of the ROPs and an iterative procedure is applied until a convergence criterion is met. The question that arises at this stage is how to employ this procedure while not restricting the flight trajectory to the x-axis and providing the capability of outlier removal through the iteration process. The possibility of doing so is discussed in the next section.

$$p_1^{\mathrm{T}}\left(\hat{T}+\delta\hat{T}\right)\delta R R p_2 = 0 \tag{17}$$

where $p_1 = (x_1, y_1, -c_1)^{\mathrm{T}}$ and $p_2 = (x_2, y_2, -c_2)^{\mathrm{T}}$ are the image coordinates of a conjugate point pair corrected for principal point offsets and camera-specific distortions, $\hat{T}$ is the 3-by-3 skew-symmetric matrix comprised from the approximate values for the translation parameters T_x, T_y, and T_z, $\delta\hat{T}$ is a 3-by-3 skew-symmetric matrix comprised from the unknown corrections δT_x, δT_y, and δT_z to the approximate translation vector, R is the approximate rotation matrix, which is defined by the approximate angles ω, ϕ, and κ derived from the assumed flight trajectory, and δR describes the unknown incremental rotation matrix, which is defined by the incremental angles $\Delta\omega$, $\Delta\phi$, and $\Delta\kappa$, that should be applied to the approximate rotation matrix to represent the true rotation between the images of the stereo-pair.

$$p_1^{\mathrm{T}}\hat{T}\delta R R p_2 + p_1^{\mathrm{T}}\delta\hat{T}\delta R R p_2 = 0 \tag{18}$$

$$\delta R = \begin{bmatrix} 1 & -\Delta\kappa & \Delta\phi \\ \Delta\kappa & 1 & -\Delta\omega \\ -\Delta\phi & \Delta\omega & 1 \end{bmatrix} \tag{19}$$

$$M_1\delta R T_3 + p_1^{\mathrm{T}} M_2 M_3 = 0$$

$$\text{where } M_1 = p_1^{\mathrm{T}}\hat{T} = \begin{bmatrix} x_1' & y_1' & z_1' \end{bmatrix},\ M_2 = \delta\hat{T}\delta R, \text{ and} \tag{20}$$

$$M_3 = R p_2 = \begin{bmatrix} x_2' & y_2' & z_2' \end{bmatrix}$$

$$\begin{bmatrix} -y_1'z_2' + z_1'y_2' & x_1'z_2' - z_1'x_2' & -x_1'y_2' + y_1'x_2' & x_1z_2' + c_1x_2' & -x_1y_2' + y_1x_2' \end{bmatrix} \begin{bmatrix} \Delta\omega \\ \Delta\phi \\ \Delta\kappa \\ \delta T_y \\ \delta T_z \end{bmatrix} \tag{21}$$

$$= \begin{bmatrix} -x_1'x_2' - y_1'y_2' - z_1'z_2' \end{bmatrix}$$

$$T_{refined} = \begin{bmatrix} T_x \\ T_y + \delta T_y \\ T_z + \delta T_z \end{bmatrix} \text{ and } R_{refined} = \delta R \bullet R \tag{22}$$

Proposed Iterative Procedure with Outlier Detection/Removal

Our objectives in this section is to extend the application of the second approach to any type of flight trajectory while providing the capability of removing outliers through the iterative parameter refinement procedure. Both objectives can be achieved by working with normalized image coordinates of the available conjugate point pairs according to epipolar geometry rather than the original image coordinates. More specifically, using the approximate estimates for the ROPs at a

given stage, one can derive normalized image coordinates according to epipolar geometry by having the image planes parallel to the base line connecting the perspective centers of the stereo pair in question while having the -axis parallel to the baseline (Cho *et al.*, 1993). Such transformation will ensure that for the normalized image coordinates, the flight trajectory is always aligned along the -axis of the image coordinate system. Moreover, the normalized image coordinates for the conjugate point pairs are expected to have a small *y-parallax* (theoretically, the *y-parallax* should be zero given the availability of noise free coordinates for true conjugate point pairs and correct ROP values), which could be used as a criterion for outlier removal.

For the relative orientation, the rotation matrix defines the rotational relationship between the right and left image coordinate systems of the stereo-pair in question; this rotation matrix can be defined as R_r^l. Using the approximate components for the translation vector (T_x, T_y, T_z), one can derive a rotation matrix R_l^n that relates the left camera coordinates system to the normalized camera coordinate system i.e., the one whose *xy*-plane is parallel to the baseline with the *x*-axis aligned along the baseline (Cho *et al.*, 1993). Using such a rotation matrix, one can derive the normalized coordinates for the conjugate point pairs as shown in Equations 23 and 24, where $c_n = 0.5(c_1 + c_2)$.

$$\begin{bmatrix} x_1^n \\ y_1^n \\ -c_n \end{bmatrix} = \lambda_1 R_l^n \begin{bmatrix} x_1 \\ y_1 \\ -c_1 \end{bmatrix} = \lambda_1 \begin{bmatrix} N_x^1 \\ N_y^1 \\ D_1 \end{bmatrix} \Rightarrow x_1^n = -c_n N_x^1 / D_1 \;\&\; y_1^n = -c_n N_y^1 / D_1 \tag{23}$$

$$\begin{bmatrix} x_2^n \\ y_2^n \\ -c_n \end{bmatrix} = \lambda_2 R_l^n R_r^l \begin{bmatrix} x_2 \\ y_2 \\ -c_2 \end{bmatrix} = \lambda_2 \begin{bmatrix} N_x^2 \\ N_y^2 \\ D_2 \end{bmatrix} \Rightarrow x_2^n = -c_n N_x^2 / D_2 \;\&\; y_2^n = -c_n N_y^2 / D_2 \tag{24}$$

The normalized image coordinates for the available conjugate point pairs can then be used in Equation 21 to derive a refined estimate for the ROPs relating the two images of the stereo-pair in question. Before using all the available conjugate point pairs, one could filter conjugate point pairs that exhibit large *y–parallax*. Moreover, additional constrains could be applied on the *x–parallax* values to remove more potential outliers. The first one is that the *x–parallax* should be always larger than zero (i.e., the corresponding object point should be below the base line). In addition, having prior information about the flying height above ground, we can impose another constraint using the approximate guess of the expected *x–parallax* as seen in Equation 25. Thus, the normalization of the image coordinates for the conjugate point pairs provide reliable information for removing potential outliers during the iterative ROP refinement process. At this stage, we should note that the estimated ROPs using the normalized image coordinates are based on the normalized camera coordinate system as defined by the orientation of the baseline in space (i.e., we will be estimating $T_{normalized}$&$R_{r_n}^{l_n}$). One can derive the updated ROPs relative to the original left camera coordinate system using Equation 26. For the iteration stopping criterion, one can use the variations of the refined ROPs between two successive iterations (i.e., we stop the iterations whenever no significant change is observed between two successive estimates of ω, ϕ, κ, and T_x, T_y, and T_z)

$$X\_parallax \approx \frac{Bc_n}{H} \tag{25}$$

where, c_n is the principal distance of the normalized images, B is the approximate length of the baseline, and H is the approximate flying height above ground.

$$\begin{aligned} R_r^l &= R_n^l R_{r_n}^{l_n} \\ T &= R_n^l T_{normalized} \end{aligned} \tag{26}$$

In summary, the proposed iterative approach starts from a linearized co-planarity model while assuming the availability of prior information regarding the system trajectory, which could be furnished by a consumer-grade GNSS/INS unit onboard the mapping platform or the predefined flight plan. For this approach, the image coordinates are normalized according to epipolar geometry to meet the underlying assumption of having the trajectory mainly aligned along the x-axis of the image coordinate system as well as providing additional constraints to help in the removal of potential outliers among available conjugate point pairs. One should note that such iterative approach can deal with stereo pairs that have any tilt angles and vertical translation component for stereo-images. However, sufficiently accurate initial approximations are required.

Experimental Results

The main objective of the experimental results is illustrating the feasibility of the proposed approaches in estimating the ROPs between the constituents of stereo-pairs captured by either multi-rotor or fixed-wing UAVs in the presence/absence of a stabilizing gimbal for the used digital cameras. In other words, the UAVs are chosen to test the ability of the proposed approaches in handling significant variation from their underlying assumptions (i.e., the images are acquired with the camera's optical axis pointing in the vertical direction and at the same flying height). The experimental datasets are picked in such a way that the tested stereo-pairs cover areas that are conducive to both high and low percentage of matching outliers. Three real datasets, which are captured by a multi-rotor DJI Phantom2 UAV with a GoPro Hero 3+ Camera (Datasets 1 and 2) and a fixed-wing PrecisionHawk UAV equipped with a Nikon J1 digital camera (Dataset 3), are used in the experimental results. The internal characteristics of the GoPro and Nikon digital cameras are estimated through a calibration procedure similar to the one proposed by He and Habib (2015). For the multi-rotor 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. For the fixed-wing UAV, the Nikon J1 camera is rigidly fixed to its body. Thanks to the stabilizing gimbal, successive images captured by the multi-rotor UAV along the same flight line can be assumed to comply with the assumptions of the two-point algorithm. On the other hand, successive images captured by the fixed-wing UAV in a given flight line are expected to show significant deviations. For both platforms, stereo-images from neighboring flight lines are also expected to show significant deviations from being almost vertical and at the same flight height. The first and second datasets are captured by the multi-rotor UAV over a building with complex roof structure and a crop field with repetitive texture, respectively. The third dataset is captured by the fixed-wing UAV over the crop field as well. The main characteristics of these datasets are described below.

- **Dataset 1** is comprised of 87 images along four strips that are captured from a flying height of roughly 20 meters with the multi-rotor UAV moving at a speed of

roughly 4 m/s. The overlap and side lap percentages for the acquired images are approximately 80 percent and 60 percent, respectively.

- **Dataset 2** includes 470 images acquired from 8 flight lines with an 8 m/s speed of the multi-rotor at a flying height of almost 15 meters. The overlap and side lap ratio among the acquired images are almost 60 percent.
- **Dataset 3** has 427 images, which are captured by the fixed-wing UAV while moving at a speed of roughly 20 m/s at a flying height of almost 55 meters. The overlap and side lap percentages for the acquired images are approximate 60 percent.

To illustrate the feasibility of the proposed approaches, we conducted several experiments with stereo-pairs that are captured along the same and neighboring flight lines. The ROP estimates from the proposed approaches are compared with those derived from manually-identified tie points (for some stereo-pairs) as well as estimated ROPs from a bundle adjustment procedure (for all the stereo-pairs within the different blocks). In this comparison, we also evaluated the performance of the Nistér five-point approach. A possible strategy for evaluating the closeness of the derived ROPs from different approaches is simply reporting the differences in the rotation and translation components of the estimated parameters. The main disadvantage of such strategy is ignoring possible correlations among the ROPs (i.e., in spite of having significant differences between the individual values for the ROPs, the accumulated effect of such parameters could be quite similar). In order to account for such correlation, the derived parameters from two different approaches are compared through a similarity analysis, which provides a single estimate in image-space units reflecting the closeness of the two ROP sets. Inspired by Habib *et al.* (2014), the proposed comparison, which is graphically illustrated in Figure 3, starts by defining a synthetic regular grid in the left image of a stereo-pair. Then, a set of 3D object points within a pre-defined elevation range, which is compatible with the Base/Height ratio for the tested stereo-pair, are simulated from the light rays connecting the perspective center and the synthetic grid vertices in the left image: see Figure 3a. The simulated 3D object points are back-projected onto the right image using the two ROP estimates that are being compared: see Figure 3b. Differences in the ROPs will cause discrepancies between the coordinates of corresponding vertices along the right image as seen in Figure 3b. The degree of similarity between such ROPs (e.g., and) can be quantitatively evaluated through the Root Mean Squared Error (RMSE) of the discrepancies between the corresponding vertices.

Figure 4 illustrates six stereo-pairs along the same flight line: Figures 4a, 4b, and 4cfor Datasets 1, 2, and 3, respectively; and from neighboring flight lines: Figures 4d, 4e, and 4f for Datasets 1, 2, and 3, respectively. For each of these stereo-pairs, we manually measured ten tie points that are used in the non-linear coplanarity model to derive the ROPs, which will be denoted here forth as the "*true* ROPs." The manually-measured tie points and automatically-derived matches from the SIFT operator and descriptor (Lowe, 2004) are then used in the proposed approaches to derive ROP estimates, which will be compared with the true ones. In addition to the ROP comparison with those derived from the coplanarity model using manually identified tie points, we also compared the estimated ROPs from the different approaches with the derived ROPs from the bundle adjustment estimates of the Exterior Orientation Parameters (EOPs). More specifically, the derived EOPs from the proposed SfM approach in He and Habib (2014) are used to estimate the ROPs between all the stereo-pairs within a given image block; such ROPs will be denoted here forth as "***Bundle Adjustment-based*** ROPs" or "***BA-based*** ROPs." One should note that the utilized SfM procedure incorporates a hybrid approach for the estimation of the ROPs for all the stereo-pairs within a block prior to the BA procedure. The hybrid approach starts with the two-point procedure whose output is used as initial values for the iterative five-point one. The SfM refines the derived ROPs from the hybrid approach through tie point tracking, rejecting incompatible ROPs, and utilizing a global bundle adjustment procedure that removes more matching outliers. Table 2 reports the number of stereo-pairs with baselines aligned along/across the flight direction for the different datasets. In order to assess the quality of the BA-based ROPs, the reconstructed surfaces from the SfM procedure are either compared with an overlapping laser scanning data or surveyed ground control points. More specifically, for Dataset 1, the SfM sparse point cloud is registered to a terrestrial laser scanner dataset using the proposed ICPatch strategy in Habib *et al.* (2010). The derived RMSE value of the discrepancies between the laser-based and SfM point clouds following the registration process is almost 3 cm, which demonstrates the accuracy of the BA-based ROPs. For Datasets 2 and 3, the datum for the SfM-based bundle adjustment is defined by ten Ground Control Points. The derived RMSE values for 18 Check Points, which are incorporated within the SfM-based bundle adjustment, for

Table 2. Number of Stereo Pairs with Baselines Aligned Along/Across the Flight Direction for the Different Datasets

Dataset	Number of stereo-pairs along the flight direction	Number of stereo-pairs across the flight direction
1	171	121
2	932	961
3	1,331	1,384

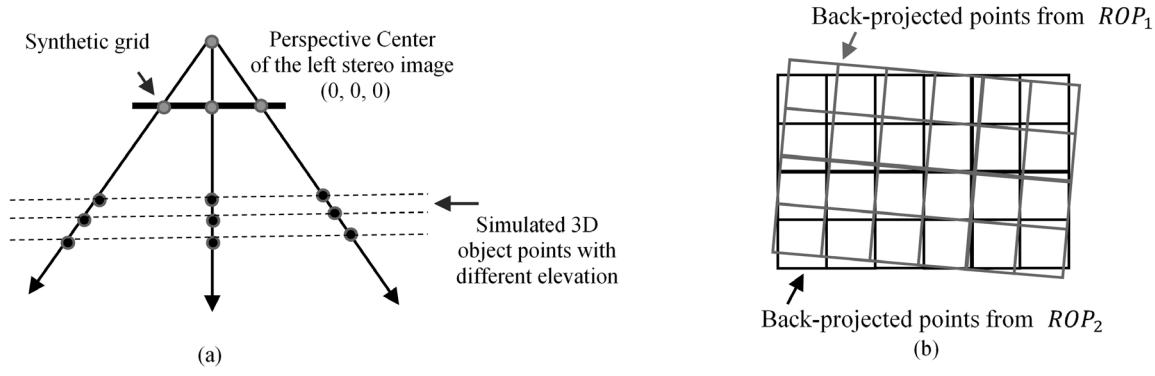

Figure 3. (a) Synthetic grid and simulated 3D object points, and (b) discrepancies between corresponding back-projected points using two ROP sets.

the two datasets are 5 cm and 4 cm, respectively. These RMSE values also confirm the accuracy of the BA-based ROPs. Therefore, the BA-based ROPs can be used as a basis for evaluating the quality of the derived ROPs from the different approaches.

As already mentioned in the Methodology Section, the proposed two-point and the Nistér five-point approaches do not require approximate values for the ROPs. For the proposed iterative-approach, on the other hand, we have to specify initial values for the ROPs. For the captured datasets by the multi-rotor UAV (Datasets 1 and 2), the initial values are estimated based on the designed flight plan. For the fixed-wing UAV (Dataset 3), we used the outcome from the two-point algorithm to initialize the iterative five-point approach. The different treatment for Dataset 3 is based on the fact that significant deviations from the flight plan should be expected when working with a rigidly-fixed camera to the UAV body.

Results and Discussion

For the stereo-based test, Table 3 reports the true ROPs, i.e., those derived from the manually-identified tie points through the non-linear coplanarity model as well as the initial values for the proposed five-point iterative approach. Since the ROPs can be only determined up to an arbitrary scale, either the T_x or T_y translation component is normalized to 1 depending on the baseline direction. Table 4 presents the differences between the estimated ROPs (i.e., those derived from the two-point, the iterative five-point, and the Nistér five-point approaches) and the true ROPs. More specifically, Rows 1, 2, and 3 for each stereo-pair present the absolute rotation and translation errors for the derived ROPs when using the manually-measured tie point coordinates in the different approaches (i.e., the two-point, iterative, and Nistér five-point approaches). Rows 4 and 5, on the other hand, show the errors associated with derived ROPs from the proposed

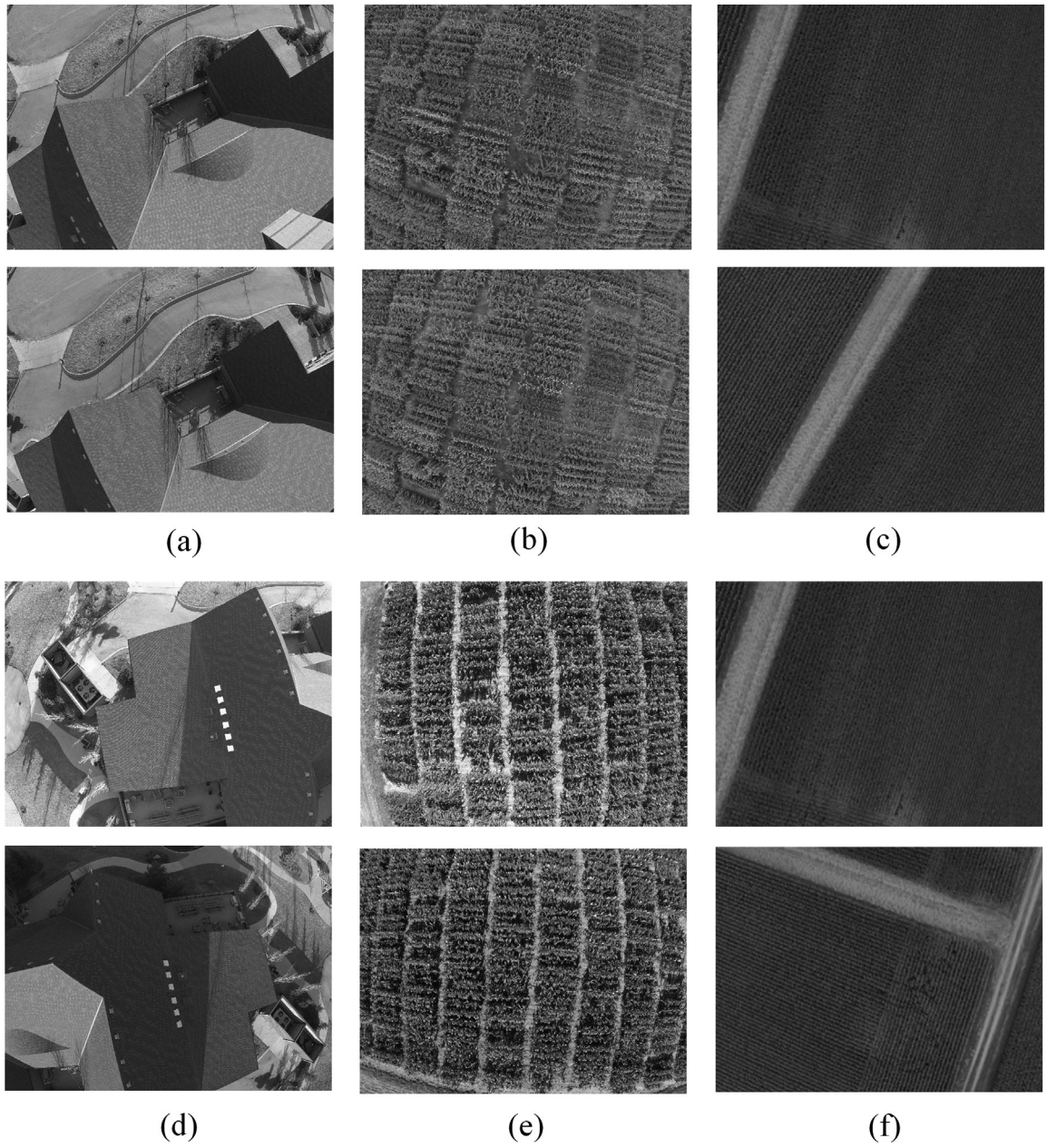

Figure 4. (a, b, and c) stereo-pairs with baseline aligned along the flight direction in Datasets 1, 2, and 3, and (d, e, f) stereo-pairs with baseline aligned across the flight direction from Datasets 1, 2, and 3.

Table 3. True ROPs from the Non-Linear Coplanarity Model while Using the Manually-Identified Tie Points and Initial Values for the Iterative Five-Point Approach for the Stereo-Based Tests (* Indicates that the Outcome from the 2-Point Approach is Used to Initialize the Iterative 5-Pint Procedure)

Test	ROP Values	ω °	ϕ°	κ°	T_x	T_y	T_z
Stereo-1-along	True (Manual)	0.25	0.14	0.59	-0.48	1.00	0.20
	Initial (Iterative)	0.00	0.00	0.00	-0.50	1.00	0.00
Stereo-1-across	True (Manual)	-13.66	4.72	179.61	1.00	0.85	-0.13
	Initial (Iterative)	0.00	0.00	180.0	1.00	0.80	0.00
Stereo-2-along	True (Manual)	0.47	0.33	-0.26	0.02	1.00	0.03
	Initial (Iterative)	0.00	0.00	0.00	0.00	1.00	0.00
Stereo-2-across	True (Manual)	0.57	0.14	178.52	1.00	0.02	0.02
	Initial (Iterative)	0.00	0.00	180.0	1.00	0.00	0.00
Stereo-3-along	True (Manual)	-1.93	11.31	12.59	-0.13	1.00	0.12
	Initial (Iterative*)	0.00	0.00	13.31	-0.29	1.00	0.00
Stereo-3-across	True (Manual)	4.87	4.38	-89.93	-1.00	0.44	0.30
	Initial (Iterative*)	0.00	0.00	-89.22	-1.00	0.38	0.00

Table 4. Comparison between the Estimated and True ROPs for the Stereo-Based Tests

Tests	Approaches	$\lvert\omega°\rvert$	$\lvert\phi°\rvert$	$\lvert\kappa°\rvert$	$\lvert T_x\rvert$ (%)	$\lvert T_y\rvert$ (%)	$\lvert T_z\rvert$ (%)	No. of initial matches	No. of inliers	No. of trials/ iterations
Stereo-1-along	2-Point + Manual	0.25	0.14	2.14	4%	0%	20%	10	7	24
	Iterative+Manual	0.02	0.03	0.04	0%	0%	0%	10	10	3
	Nistér + Mannual	0.21	0.28	0.15	1%	0%	0%	10	9	31
	2-Point + SIFT	0.25	0.14	0.14	1%	0%	20%	4,894	4,344	58
	Iterative + SIFT	0.21	0.15	0.66	0%	0%	0%	4,894	4,811	8
	Nistér + SIFT	0.42	0.21	1.34	12%	0%	7%	4,894	4,794	6
Stereo-1-across	2-Point + Manual	13.66	4.72	7.05	0%	16%	13%	10	7	24
	Iterative+ Manual	0.02	0.02	0.01	0%	0%	1%	10	10	4
	Nistér + Mannual	0.27	0.19	0.24	0%	0%	0%	10	10	18
	2-Point + SIFT	13.66	4.72	7.27	0%	16%	13%	1,822	742	104
	Iterative + SIFT	1.06	3.30	1.13	0%	0%	0%	1,822	1,683	7
	Nistér + SIFT	0.27	1.83	2.12	0%	1%	1%	1,822	1,525	16
Stereo-2-along	2-Point + Manual	0.47	0.33	0.98	2%	0%	3%	10	6	43
	Iterative+ Manual	0.00	0.00	0.01	1%	0%	0%	10	10	4
	Nistér + Mannual	0.31	0.13	0.26	1%	0%	0%	10	9	32
	2-Point + SIFT	0.47	0.33	3.42	2%	0%	3%	554	55	1,732
	Iterative + SIFT	0.32	0.28	0.01	1%	0%	1%	554	78	10
	Nistér + SIFT	Stopped after 100,000 iterations								
Stereo-2-across	2-Point + Manual	0.57	0.14	0.36	0%	2%	2%	10	6	27
	Iterative+ Manual	0.04	0.02	0.05	0%	1%	2%	10	10	5
	Nistér + Mannual	0.27	0.18	0.15	0%	1%	1%	10	8	53
	2-Point + SIFT	0.57	0.14	2.08	0%	3%	2%	633	42	2,571
	Iterative + SIFT	3.57	3.89	2.90	0%	1%	2%	633	62	33
	Nistér + SIFT	Stopped after 100,000 iterations								
Stereo-3-along	2-Point + Manual	1.93	11.31	0.72	16%	0%	12%	10	5	28
	Iterative+Manual	0.20	0.02	0.18	2%	0%	1%	10	10	6
	Nistér + Mannual	0.19	0.13	0.22	2%	0%	2%	10	8	47
	2-Point + SIFT	1.93	11.31	0.64	16%	0%	12%	2,697	473	932
	Iterative + SIFT	0.82	0.50	1.44	2%	0%	1%	2,697	1,104	4
	Nistér + SIFT	1.64	0.92	0.11	1%	0%	1%	2,697	1,072	208
Stereo-3-across	2-Point + Manual	4.87	4.38	0.34	0%	5%	30%	10	6	13
	Iterative+ Manual	0.26	0.16	0.13	0%	0%	0%	10	10	5
	Nistér + Mannual	0.27	0.22	0.17	2%	0%	1%	10	9	26
	2-Point + SIFT	4.87	4.38	0.71	0%	6%	30%	2,188	260	2,561
	Iterative + SIFT	0.15	1.32	1.49	0%	0%	1%	2,188	701	6
	Nistér + SIFT	20.99	1.85	182.8	200%	24%	0.01	2,188	632	453

approaches while incorporating the automatically-identified conjugate point pairs through the SIFT operator and descriptor. Finally, Row 6 provides the ROP errors for the Nistér five-point approach. In Table 4, one should note that the absolute translation errors are presented as an error percentage since the translation components are normalized according to the baseline direction. Table 4 also reports the number of input matches (manually-based or SIFT-based tie points), the identified conjugate point pairs by the ROP procedure, and the number of trials/iterations performed by the different approaches in Columns 9, 10, and to 11, respectively. The RMSE similarity values, which describe the closeness of the different ROPs to the true ones, are illustrated in Table 5.

For the block-based tests, Table 6 reports the statistics of the differences between the estimated ROPs from the two-point or the iterative five-point approaches and the BA-based ROPs. More specifically, Rows 4 through 9 present the RMSE values for the rotation and translation differences for the three datasets while considering different configurations (i.e., stereo-pairs with baseline aligned along/across the flight direction). Rows 10 through 15, on the other hand, present the maximum rotation and translation errors. Finally, Table 7 introduces the statistics of the RMSE similarity values, which are derived from the proposed similarity analysis, for all the stereo-pairs in the three datasets.

Based on the reported results for the stereo-based and block-based tests, the following observations can be made:

- The absolute difference and similarity-based RMSE values give compatible judgment regarding the closeness of the estimated ROPs from the different approaches and the true or BA-based ROPs. This can be clearly seen in the reported differences and RMSE similarity values in Tables 4 through 6 and 5 through 7, respectively. This is an indication that there is no correlation among the estimated ROPs (i.e., sufficiently strong geometry of tie points is inherent within all the involved stereo-pairs).
- The proposed approaches are capable of handling expected variations from the assumed flight plan and stereo-configurations when dealing with stereo-pairs in different flight lines and/or captured stereo-pairs along the same flight line by digital cameras that are rigidly-fixed to the UAV body which is the case for Dataset 3.
- Comparing the two-point, the iterative five-point, and the Nistér five-point approaches, the iterative five-point approach resulted in the closest ROPs to the true and BA-based ROPs.
- When dealing with stereo-pairs that exhibit significant variations from being captured at the same flying height with the camera in a nadir looking configuration, the

Table 5. RMSE Similarity Values between the Estimated and True ROPs for the Stereo-Based Tests

ROP_1 & ROP_2	RMSE for Stereo-1-along (pixel)	RMSE for Stereo-1-across (pixel)	RMSE for Stereo-2-along (pixel)	RMSE for Stereo-2-across (pixel)	RMSE for Stereo-3-along (pixel)	RMSE for Stereo-3-across (pixel)
True & 2-Point + Manual	42.13	215.13	29.81	30.09	341.13	243.07
True & Iterative + Manual	1.01	1.07	2.67	7.21	16.02	8.31
True & Nistér + Manual	9.82	11.35	15.86	16.97	33.52	16.23
True & 2-Point + SIFT	21.79	219.68	51.14	67.33	320.11	251.08
True & Iterative + SIFT	12.30	79.52	17.42	82.51	54.61	37.11
True & Nistér + SIFT	32.07	63.73	N/A	N/A	65.47	N/A

Table 6. Comparison between the Estimated and BA-Based ROPs for the block-based tests

Dataset	Dataset 1				Dataset 2				Dataset 3			
Approach	2-Point		Iterative		2-Point		Iterative		2-Point		Iterative	
No. of Stereos	171-along	121-across	171-along	121-across	932-along	961-across	932-along	961-across	1,331-along	1,384-across	1,331-along	1,384-across
RMSE_$\Delta\omega$(°)	0.73	5.18	0.21	0.51	0.12	2.24	0.52	0.77	5.32	6.39	0.70	1.17
RMSE_$\Delta\phi$(°)	0.65	3.43	0.64	0.48	0.06	3.46	0.18	0.50	6.61	7.77	0.82	0.96
RMSE_$\Delta\kappa$(°)	0.69	4.25	0.33	0.57	0.83	1.35	0.21	0.36	5.90	5.30	0.31	0.66
RMSE_ΔT_x(%)	1%	0%	1%	0%	2%	0%	0%	0%	16%	9%	6%	3%
RMSE_ΔT_y(%)	0%	8%	0%	2%	0%	8%	0%	2%	0%	21%	3%	7%
RMSE_ΔT_z(%)	11%	19%	0%	1%	3%	4%	1%	0%	13%	15%	4%	8%
Max_$\lvert\Delta\omega\rvert$(°)	2.63	13.66	1.21	2.97	0.30	3.30	1.46	3.07	12.33	13.84	1.71	2.31
Max_$\lvert\Delta\phi\rvert$(°)	3.79	4.72	0.79	2.05	0.18	3.85	0.98	1.81	15.22	19.04	2.23	1.79
Max_$\lvert\Delta\kappa\rvert$(°)	2.13	7.56	0.88	1.13	3.93	5.30	0.97	1.94	18.90	161.9	1.10	2.09
Max_ $\lvert\Delta T_x\rvert$(%)	11%	0%	5%	0%	8%	0%	2%	0%	47%	15%	1%	4%
Max_ $\lvert\Delta T_y\rvert$(%)	0%	12%	0%	6%	0%	17%	0%	9%	0%	51%	0%	5%
Max_ $\lvert\Delta T_z\rvert$(%)	21%	17%	6%	4%	5%	9%	5%	3%	32%	42%	1%	3%

Table 7. RMSE Similarity Values between the Estimated and BA-Based ROPs for the Block-Based Tests

ROP_1 & ROP_2	RMSE values for Dataset 1 (pixel)			RMSE values for Dataset 2 (pixel)			RMSE values for Dataset 3 (pixel)		
	Mean	Std	Max	Mean	Std	Max	Mean	Std	Max
BA-based & 2-Point + SIFT (along flight line)	10.01	19.83	136.27	8.50	14.61	87.37	115.06	125.79	346.62
BA-based & 2-Point + SIFT (across flight line)	82.31	85.69	361.08	65.34	68.27	225.16	134.57	166.13	392.15
BA-based & Iterative + SIFT (along flight lien)	5.28	7.11	73.71	4.44	6.61	69.68	21.54	31.01	87.31
BA-based & Iterative + SIFT (across flight line)	9.08	14.35	112.84	9.16	24.96	139.68	33.47	46.73	144.28

two-point approach results in larger differences from the true and BA-based ROPs (refer to the highlighted cells in Table 4 where the maximum ω and ϕ rotation angles are 13.66° and 11.31°, respectively, while the largest normalized T_z component is almost 0.30). The impact of such variations is clearly visible in terms of low percentage of accepted matches in the two-point approach (even when dealing with manually-measured tie points).

- The iterative and Nistér five-point approaches exhibit similar performance when dealing with the stereo-pairs that have high percentage of correct matches (e.g., those in Dataset 1). However, the Nistér five-point approach performs poorly when dealing with stereo-pairs contaminated by high percentage of matching outliers. Table 4 shows that for Datasets 2 and 3, the Nistér five-point approach fails when the percentage of outliers reaches almost 90 percent and 60 percent, respectively.
- A hybrid approach that starts with the two-point approach and uses the estimated ROPs as initial values for the iterative five-point approach proved to be a good strategy when the stereo-pairs exhibit significant variations from the design plan, which is usually the case for UAV-based flights (e.g., wind conditions and the light-weight of the UAVs will lead to significant variations from the design plan and deviations from the assumption of the two-point algorithm).

Conclusions and Recommendations for Future Work

This paper presents two approaches for ROP estimation while dealing with UAV-based imagery. Both approaches take advantage of prior information regarding the flight trajectory, which can be derived from the designed mission plan and/or geo-referencing information from an onboard GNSS/INS unit. The first approach assumes that the UAV platform is moving at 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). Starting from such assumptions, the nine elements of the Essential matrix relating conjugate points in overlapping images are reduced to four with an additional constraint among them. Given that the relative orientation can be only established up to an arbitrary scale, the four elements of the Essential matrix can be derived using a minimum of two point correspondences. This approach can be incorporated within a RANSAC framework to remove potential outliers among the initial conjugate point pairs. Thanks to the fact that it only requires a minimum of two parameters, fewer RANSAC trials will be needed when compared to existing five and eight point algorithms. The second approach starts with a linearization process of the co-planarity constraint starting from prior information regarding the ROPs relating the stereo-pair in question to derive a set of linear equations in the unknown corrections to the approximate ROP values. A minimum of five conjugate pairs are required to derive an estimate of the ROPs through an iterative procedure until a convergence criterion is achieved. Thanks to the built-in normalization procedure according to epipolar geometry, this approach can remove matching outliers by imposing constraints on the *xy–parallax* values. A common characteristics of both approaches is their ability to use more than the minimum number of required conjugate point pairs to provide an LSA-based estimate of the ROPs. Experimental results have shown that both procedures are capable of providing accurate ROP estimates in the presence of high percentage of matching outliers, which could be the result of dealing with images that exhibit repetitive pattern. Although the iterative approach has shown better performance when compared with the two-point approach, the latter does not make any restrictions regarding the heading difference as well as the planimetric motion of imaging platform. However, this comes at the price that the two-point approach cannot tolerate significant variations from having vertical images that are captured at the same flying height. These characteristics lead to the fact that both approaches can be integrated in a hybrid strategy where the two-point procedure is utilized to provide initial ROP estimates, which could be then refined through the implementation of the iterative approach. Thus, the hybrid approach can deal with situations where prior information regarding the flight trajectory is not accurate; this should be expected when operating a light-weight UAV in relatively windy conditions. The evaluation of the integrated process on more UAV platforms will be the focus of future research. Moreover, since the proposed two-point approach imposes strict restrictions on ROPs to be estimated, other approaches that require less assumptions regarding the orientation of the utilized platform will be investigated for the automated relative orientation recovery of UAV-based imagery.

References

Cho, W., T. Schenk, and M. Madani, 1993. Resampling digital imagery to epipolar geometry, *International Archives of Photogrammetry and Remote Sensing*, 29:404–404.

Cramer, M., D. Stallmann, and N. Haala, 2000. Direct georeferencing using GPS/inertial exterior orientations for photogrammetric applications, *International Archives of Photogrammetry and Remote Sensing*, 33:198–205.

Faugeras, O.D., and S. Maybank, 1990. Motion from point matches: Multiplicity of solutions, *International Journal of Computer Vision*, 4:225–246.

Fischler, M.A., and R.C. Bolles, 1981. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography, *Communications of the ACM*, 24:381–395.

Fraser, C.S., 1997. Digital camera self-calibration, *ISPRS Journal of Photogrammetry and Remote sensing*, 52:149–159.

Fraundorfer, F., P. Tanskanen, and M. Pollefeys, 2010. A minimal case solution to the calibrated relative pose problem for the case of two known orientation angles, *Proceedings of Computer Vision–ECCV 2010 Conference*, Springer, pp. 269–282.

Gallagher, A.C., 2005. Using vanishing points to correct camera rotation in images, *Proceedings of the 2nd Canadian Conference on Computer and Robot Vision*, 2005. IEEE, pp. 460–467.

Habib, A., I. Detchev, and K. Bang, 2010. A comparative analysis of two approaches for multiple-surface registration of irregular point clouds, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 38(Part 1):61–66.

Habib, A., I. Detcheve, ad E. Kwak, 2014. Stability analysis for a multi-camera photogrammetric system, *Sensors*, 14(8):15084–15112.

Habib, A.F., and M.F. Morgan, 2003. Automatic calibration of low-cost digital cameras, *Optical Engineering*, 42:948–955.

Habib, A., and D. Kelley, 2001. Automatic relative orientation of large scale imagery over urban areas using modified iterated Hough transform, *ISPRS journal of Photogrammetry and Remote Sensing*, 56:29–41.

Habib, A., and M. Mazaheri, 2015. Quaternion-based solutions for the single photo resection problem, *Photogrammetric Engineering & Remote Sensing*, 81(2):209–217.

Hartley, R., 1997. In defense of the eight-point algorithm, *IEEE Transactions on Pattern Analysis and Machine Intelligence*,19:580–593.

Hartley, R., and A. Zisserman, 2003. *Multiple View Geometry in Computer Vision*, Second edition, Cambridge University Press.

He, F., and A. Habib, 2015. Target-based and feature-based calibration of low-cost digital cameras with large field-of-view, *Proceedings of ASPRS 2015 Annual Conference*, Tampa, Florida.

He, F., and A. Habib, 2014. Linear approach for initial recovery of the exterior orientation parameters of randomly captured images by low-cost mobile mapping systems, *ISPRS-International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-1:149–154.

He, F., A. Habib, and A. Al-Rawabdehb, 2015. Planar constraints for an improved UAV-image-based dense point cloud generation, *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-1/W4:269–274.

Heipke, C., 1997. Automation of interior, relative, and absolute orientation, *ISPRS Journal of Photogrammetry and Remote Sensing*, 52:1–19.

Horn, B.K., 1990a. Relative orientation, *International Journal of Computer Vision*, 4:59–78.

Horn, B.K., 1990b. Recovering baseline and orientation from essential matrix, *Journal of the Optical Society of America*, pp. 110.

Huang, T.S., and A.N. Netravali, 1994. Motion and structure from feature correspondences: A review, *Proceedings of the IEEE*, 82:252–268.

Kalantari, M., A. Hashemi, F. Jung, and J.-P. Guédon, 2011. A new solution to the relative orientation problem using only 3 points and the vertical direction, *Journal of Mathematical Imaging and Vision*, 39:259–268.

Lari, Z., A. Al-Rawabdeh, F. He, A. Habib, and N. El-Sheimya, 2015. Region-based 3D surface reconstruction using images acquired by low-cost unmanned aerial systems, *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-1/W4:167–173.

Longuet-Higgins, H.C., 1987. A computer algorithm for reconstructing a scene from two projections, *Readings in Computer Vision: Issues, Problems, Principles, and Paradigms* (M.A. Fischler and O. Firschein, editors), pp. 61–62.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision*, 60:91–110.

Mikhail, E.M., J.S. Bethel, and J.C. McGlone, 2001. *Introduction to Modern Photogrammetry*, John Wiley & Sons, Inc.

Naroditsky, O., X.S. Zhou, J. Gallier, S. Roumeliotis, and K. Daniilidis, 2012. Two efficient solutions for visual odometry using directional correspondence, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34:818–824.

Nistér, D., 2004. An efficient solution to the five-point relative pose problem, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:756–770.

Ortin, D., and J.M.M. Montiel, 2001. Indoor robot motion based on monocular images, *Robotica*, 19:331–342.

Philip, J., 1998. Critical point configurations of the 5-, 6-, 7-, and 8-point algorithms for relative orientation, *Technical Report TRITA-MAT-1998-MA-13*, Department of Mathematics, Royal Institute of Technology, Stockholm, Sweden.

Philip, J., 1996. A non-iterative algorithm for determining all essential matrices corresponding to five point pairs, *The Photogrammetric Record*, 15:589–599.

Remondino, F., and S. El-Hakim, 2006. Image-based 3D modelling: A review, *The Photogrammetric Record*, 21:269–291.

Robertson, D.P., and R. Cipolla, 2004. An image-based system for urban navigation, *Proceedings of British Machine Vision Conference (BMVC)*, pp. 1–10.

Scaramuzza, D., 2011a. Performance evaluation of 1-point-RANSAC visual odometry, *Journal of Field Robotics*, 28:792–811.

Scaramuzza, D., 2011b. 1-point-RANSAC structure from motion for vehicle-mounted cameras by exploiting non-holonomic constraints, *International Journal of Computer Vision*, 95:74–85.

Scaramuzza, D., F. Fraundorfer, and R. Siegwart, 2009. Real-time monocular visual odometry for on-road vehicles with 1-point RANSAC, *Proceedings of 2009 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4293–4299.

Skaloud, J., 2002. Direct georeferencing in aerial photogrammetric mapping, *Photogrammetric Engineering & Remote Sensing*, 68:207–210.

Stewenius, H., C. Engels, and D. Nistér, 2006. Recent developments on direct relative orientation, *ISPRS Journal of Photogrammetry and Remote Sensing*, 60:284–294.

Tang, L., and C. Heipke, 1996. Automatic relative orientation of aerial images, *Photogrammetric Engineering & Remote Sensing*, 62(4):47–55.

Triggs, B., 2000. Routines for relative pose of two calibrated cameras from 5 points, Technical Report, URL: *https://hal.inria.fr/inria-00548287/document*, France (last data accessed: 20 September 2016).

Troiani, C., A. Martinelli, C. Laugier, and D. Scaramuzza, 2014. 2-point-based outlier rejection for camera-imu systems with applications to micro aerial vehicles, *Proceedings of 2014 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5530–5536.

Viéville, T., E. Clergue, and P.D.S. Facao, 1993. Computation of ego-motion and structure from visual and inertial sensors using the vertical cue, *Proceedings of Fourth International Conference on Computer Vision*, pp. 591–598.

Zhang, Y., X. Huang, X. Hu, F. Wan, and L. Lin, 2011. Direct relative orientation with four independent constraints, *ISPRS Journal of Photogrammetry and Remote Sensing*, 66:809–817.

(Received 17 November 2015; accepted 06 June 2016; final version 20 June 2016)