

Performance Evaluation of Alternative Relative Orientation Procedures for UAV-based Imagery with Prior Flight Trajectory Information

F. He ^{a,*}, A. Habib ^a

^a Digital Photogrammetry Research Group
Lyles School of Civil Engineering
Purdue University, 550 Stadium Mall Dr., West Lafayette, IN 47907, USA
– (he270, ahabib)@purdue.edu

Commission III, WG III/1

KEY WORDS: Relative Orientation, Essential Matrix, Closed-form Solution, UAVs

ABSTRACT:

Thanks to recent advances at the hardware (e.g., emergence of reliable platforms at low cost) and software (e.g., automated identification of conjugate points in overlapping images) levels, UAV-based 3D reconstruction has been widely used in various applications. However, mitigating the impact of outliers in automatically matched points in UAV imagery, especially when dealing with scenes that has poor and/or repetitive texture, remains to be a challenging task. In spite of the fact that existing literature has already demonstrated that incorporating prior motion information can play an important role in increasing the reliability of the matching process, there is a lack of methodologies that are mainly suited for UAV imagery. Assuming the availability of prior information regarding the trajectory of a UAV-platform, this paper presents a two-point approach for reliable estimation of Relative Orientation Parameters (ROPs) of UAV-based images. This approach is based on the assumption that the UAV platform is moving at a constant flying height while maintaining the camera in a nadir-looking orientation. For this flight scenario, a closed-form solution that can be derived using a minimum of two pairs of conjugate points is established. In order to evaluate the performance of the proposed approach, experimental tests using real stereo-pairs acquired from different UAV platforms have been conducted. The derived results from the comparative performance analysis against the Nistér five-point approach demonstrate that the proposed two-point approach is capable of providing reliable estimate of the ROPs from UAV-based imagery in the presence of poor and/or repetitive texture with high percentage of matching outliers.

1. INTRODUCTION

Automated relative orientation, which defines the position and orientation of one image relative to another one, has been investigated within both the photogrammetric and computer vision research communities (Habib and Kelley, 2001; Heipke, 1997; Zhang et al., 2011). In general scenarios, the IOPs of the utilized camera to capture such images are assumed to be known for the estimation of ROPs. Therefore, for a given stereo-pair, ROP estimation involves the derivation of 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 implemented using a minimum of five conjugate points. However, due to the nonlinear nature of the co-planarity model, approximate values for the unknowns have to be available. In order to resolve such problem, 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. However, 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 has been 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 demonstrate that with image coordinate normalization, the performance of the eight-point algorithm is almost at the same quality level as the iterative non-linear algorithm. In the meantime, several five-point approaches have been proposed as alternatives to the eight-point approach (Faugeras and Maybank, 1990; Philip, 1998). The most efficient five-point algorithm is the one proposed by Nistér, (2004). Compared to the eight-point approach, the five-point approaches take into account the inherent constraints among the elements of the Essential Matrix.

Instead of using the conventional five/eight-point approach, several research efforts have been geared towards ROP recovery while taking advantage of prior information regarding the system trajectory during data acquisition. To date, assuming the knowledge of some parameters, several approaches, which are mainly initiated by the mobile robotics community, have been introduced to derive reliable ROP estimates. For example, a two-point approach has been introduced by Troiani et al. (2014) for the estimation of translation components while relying on three known rotation angles. Also, a three-point approach has been proposed by Fraundorfer et al. (2010) while assuming that the involved stereo-images share a common axis of rotation, which

* Corresponding author.

is denoted as reference direction (Viéville et al., 1993). Compared to the five/eight algorithms, one can argue that these approaches can be more advantageous, since they require fewer conjugate pairs for ROP estimation. Such requirement would reduce the number of RANSAC trials for outlier removal.

Current ROP recovery while considering prior information about the system trajectory has been mainly focusing on indoor and outdoor terrestrial mobile mapping systems. However, the manipulation of stereo-images captured by a UAV-based mapping platform has not been addressed. In this paper, we investigate the estimation of relative orientation parameters for UAV-based images using alternative approaches. More specifically, a novel two-point closed-form solution, which takes advantage of prior information regarding the flight trajectory, are initially presented. Then a comparative analysis of the derived ROPs from the proposed two-point approach as well as the conventional five approach is conducted.

The remainder part of the paper starts with the conceptual basis for the essential matrix. Then, the mathematical details for the proposed two-point approach is presented. Afterwards, experimental results using real datasets are presented. Finally, drawn conclusions and recommendations for future work are introduced.

2. CONCEPTUAL BASIS FOR THE ESSENTIAL MATRIX

The conceptual basis for the Essential matrix is based on the coplanarity constraint, which has been used in the photogrammetric research community for decades. As shown in Figure 1, the coplanarity constraint mathematically describes the fact that an object point P , the corresponding image points, and the two perspective centres O_1 and O_2 of a stereo-pair must lie on the same plane (Equation 1).

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

In this equation, p_1 and p_2 are two corresponding points, where $p = (x, y, -c)^T$ represents the image coordinates corrected for the principal point offset and camera-specific distortions. 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_y, T_z) . In the meantime, the cross product in Equation 1 can be simplified using the skew-symmetric matrix \hat{T} , which converts the cross product of the two vectors to a matrix-vector multiplication, in Equation 2. As a result, the expression for the Essential Matrix can be derived.

$$p_1^T \hat{T} R p_2 = p_1^T E p_2 = 0 \quad (2)$$

Where,

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

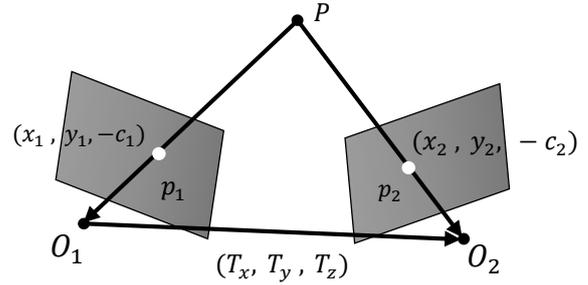


Figure 1. The co-planarity model relating stereo-images

Looking into the expression for the Essential Matrix as shown in Equation 2, one can note that the nine elements of the Essential matrix are defined by the five elements of the ROPs (three rotation angles and two translation components). Therefore, four additional constraints have to be imposed on the nine elements of the Essential matrix E . Such constraints can be explained as follows:

1. The Essential matrix has rank two. Therefore, its determinant has to be zero as shown in Equation 3.

$$\det(E) = 0 \quad (3)$$

2. The Essential matrix has two equal non-zero singular values. Therefore, two independent equations on the nine unknown parameters can be deduced from the trace constraint as presented in Equation 4.

$$EE^T E - \frac{1}{2} \text{trace}(EE^T) E = 0 \quad (4)$$

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

3. TWO-POINT APPROACH

The proposed two-point 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, two geometric constraints can be used to reduce and simplify the elements of the Essential matrix. The two geometric constraints can be explained as follows:

1. The rotation angles ω and ϕ are assumed to be zero, since a nadir-looking camera is utilized for data acquisition.
2. The T_z translation component is assumed to be zero, since the utilized UAV platform is moving at a constant flying height.

Considering these two geometric constraints, the rotation matrix R can be defined by the rotation angle κ (i.e., heading), and the translation \vec{T} can be defined by the two translation components $(T_x$ and $T_y)$ describing of the horizontal planar motion of the UAV platform. Therefore, the simplified Essential matrix can be established as the form presented in Equation 5, where L_1, L_2, L_3 , and L_4 are used to denote the four unknown parameters of the Essential matrix E .

$$\begin{aligned}
 E &= \hat{T}R \\
 &= \begin{bmatrix} 0 & 0 & T_y \\ 0 & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix} \begin{bmatrix} \cos \kappa & -\sin \kappa & 0 \\ \sin \kappa & \cos \kappa & 0 \\ 0 & 0 & 1 \end{bmatrix} \\
 &= \begin{bmatrix} 0 & 0 & T_y \\ 0 & 0 & -T_x \\ -T_y \cos \kappa + T_x \sin \kappa & T_y \sin \kappa + T_x \cos \kappa & 0 \end{bmatrix} \quad (5) \\
 &= \begin{bmatrix} 0 & 0 & L_1 \\ 0 & 0 & L_2 \\ L_3 & L_4 & 0 \end{bmatrix}
 \end{aligned}$$

As can be seen in Equation 5, 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 Essential matrix. A closer inspection of the relationships between (L_1, L_2, L_3, L_4) and (T_x, T_y, κ), one can introduce the constraint in Equation 6.

$$L_1^2 + L_2^2 - L_3^2 - L_4^2 = 0 \quad (6)$$

In addition, considering the fact that these parameters can be only determined up to an arbitrary scale, two conjugate point pairs should be sufficient for deriving the simplified Essential matrix. In this research, a closed-form solution, which is similar to the one proposed by Nistér, (2004) for the five-point approach, is adopted for the estimation of the Essential Matrix. More specifically, a second-order polynomial, which provides two possible estimates for the simplified Essential Matrix as shown in Equation 5, can be established. One should note that, since a total of four possible solutions of the rotation matrix R and translation vector \vec{T} can be recovered from a single Essential matrix (Horn, 1990), up to eight solutions for R and \vec{T} can be derived from the proposed two-point approach. In order to identify the valid Essential matrix among the available solutions, in this research, two additional constraints are utilized as follows:

1. The light rays connecting a derived object point and perspective centres should be on the same side of the baseline.
2. The derived object points should be in front of the camera.

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 can be 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. One should also note that, similar to the conventional five/eight point approach, the proposed two-point approach can be incorporated within a RANSAC framework for outlier removal.

4. EXPERIMENTAL RESULTS

The main objective of the experimental results is providing a comparative analysis of the derived ROPs from the proposed two-point and the Nistér five point approaches. Meanwhile, in order to evaluate the feasibility of the proposed two point approach in handling significant variations from the 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 utilized experimental datasets are captured by either multi-rotor or fixed-wing UAVs in the presence/absence of

a stabilizing gimbal for the used digital cameras. To be more specific, three tests are performed on image stereo-pairs that are captured by a multi-rotor DJI Phantom2 UAV with a GoPro Hero 3+ Camera (Tests 1 and 2) and a fixed-wing PrecisionHawk UAV equipped with a Nikon J1 digital camera (Test 3). 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, and no camera stabilizer system is utilized. One should note that the three experimental tests are proposed in such a way that the tested UAV images cover areas that are conducive to both high and low percentage of matching outliers. The main characteristics for the three tests are described below.

Test 1 includes two stereo-pairs that are captured by the multi-rotor UAV over a building with complex roof structure. The flying height of the UAV is roughly 20 meters, and the flying speed is roughly 4 m/s. More specifically, for the two stereo-pairs, one is along the same flight line, and the other one is from the neighboring flight lines. The overlap and side lap percentages for the acquired images are approximately 80% and 60%, respectively.

Test 2 contains two stereo-pairs that are captured by the multi-rotor UAV with an 8 m/s speed at a flying height of almost 15 meters over a crop field with repetitive patterns. The overlap and side lap ratio for the acquired image stereo-pairs are almost 60%.

Test 3 has two stereo-pairs, 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 involved test site is the same crop field for test 2. The overlap and side lap percentages for the acquired images are approximately 60%.

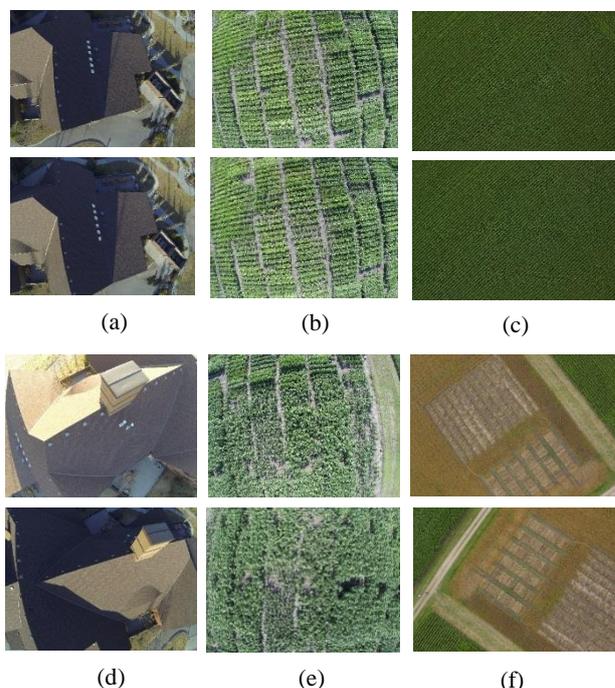


Figure 2. (a – c) Stereo-pairs with baseline aligned along the flight direction in Datasets 1 – 3, and (d – f) stereo-pairs with baseline aligned across the flight direction from Datasets 1 – 3.

Figure 2 illustrates the six utilized stereo-pairs for Tests 1, 2, and 3. Specifically, Figures 2(a), 2(b), and 2(c) illustrate the three stereo-pairs along the same flight line for Tests 1, 2, and 3,

respectively, while Figures 2(d), 2(e), and 2(f) show the other three stereo-pairs from the neighboring flight lines for Tests 1, 2, and 3. For each of these stereo-pairs, 10 tie points are manually measured and used in the non-linear coplanarity model to derive the ROPs, which will be denoted here forth as the “true ROPs”. Then, we use the automatically-derived matches from the SIFT operator and descriptor (Lowe, 2004) in both the proposed two-point and the Nistér five point approaches with the RANSAC framework to derive ROP estimates, which will be compared with the true ones.

4.1 Results and Discussions

Table 1 presents the differences between the estimated ROPs (i.e., those derived from the two-point and the Nistér five-point approaches) and the true ROPs. More specifically, the errors associated with derived ROPs from the two adopted approaches while incorporating the automatically-identified conjugate point pairs are shown in Rows 1 and 2 for test, respectively. Looking into the results in Table 1, 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 1 also reports the number of input matches (i.e., SIFT-based tie points), the identified conjugate point pairs by the ROP procedure, and the trials/iterations performed by the proposed two-point and the Nistér five-point approaches in Columns 9, 10, and to 11, respectively.

Based on the reported results for Tests 1, 2, and 3, the following observations can be made:

1. For all the three tests, the Nistér five-point approach resulted in closer ROPs to the manual-based true values when comparing to the proposed two-point approach. This is expected since for the proposed approach, we are not allowing the rotation angles ω and ϕ as well as the T_z translation component to partially absorb the impact of the noise in the image coordinates.
2. In spite of its sensitivity to deviations from having vertical images, the proposed two-point approach is still capable of handling expected variations from the assumed flight plan and stereo-configurations when dealing with stereo-pairs captured by digital cameras that are mounted on either multi-rotor or fixed-wing UAVs in the presence/absence of a camera stabilizer system.
3. The proposed two-point and the Nistér five-point approaches exhibit similar performance when dealing with the stereo-pairs that have high percentage of correct matches (e.g., those in Test 1). However, the Nistér five-point approach performs poorly when dealing with stereo-pairs contaminated by high percentage of matching outliers (as can be seen in Table 1, the Nistér five-point approach fails when the percentage of outliers reaches almost 90% – this is the case for Tests 2 where images exhibit repetitive pattern).

Test	Approaches	$ \Delta\omega^0 $	$ \Delta\kappa^0 $	$ \Delta\kappa^0 $	$ \Delta T_x $ (%)	$ \Delta T_y $ (%)	$ \Delta T_z $ (%)	No. of initial matches	No. of inliers	No. of trials/iterations
Test-1-along	2-Point + SIFT	0.57	0.21	0.86	1%	0%	3%	3,761	2,615	47
	Nistér + SIFT	0.39	0.46	0.53	2%	0%	2%	3,761	3,249	11
Test-1-across	2-Point + SIFT	2.31	3.15	1.16	0%	8%	5%	1,675	639	72
	Nistér + SIFT	0.60	1.21	0.97	0%	2%	1%	1,675	1,355	26
Test-2-along	2-Point + SIFT	0.33	0.19	1.06	1%	0%	2%	938	113	933
	Nistér + SIFT				Stopped after 100,000 iterations					
Test-2-across	2-Point + SIFT	1.38	2.01	1.16	0%	1%	3%	596	62	1,785
	Nistér + SIFT				Stopped after 100,000 iterations					
Test-3-along	2-Point + SIFT	2.09	4.39	2.55	5%	0%	3%	762	95	868
	Nistér + SIFT				Stopped after 100,000 iterations					
Test-3-across	2-Point + SIFT	5.22	3.63	1.79	0%	4%	3%	3,164	1,658	304
	Nistér + SIFT	0.91	0.55	1.19	0%	2%	1%	3,164	2,481	239

Table 1. Comparison between the estimated and true ROPs for Tests 1, 2, and 3

5. CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK

This paper presents a two-point approach for reliable relative orientation recovery of UAV-based images. Different from the conventional closed-form solutions, the proposed approach takes 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. In order to evaluate the feasibility of the proposed approach, three experimental tests, which utilize real image stereo-pairs acquired from either multi-rotor or fixed-wing UAVs, are conducted for a comparative analysis between the proposed two-point and the Nistér five point approaches. The derived experimental results demonstrate that the proposed two-point approach has superior performance when dealing with stereo-images in the presence of a high percentage of outliers.

It is important to note that, in this research, the proposed two-point approach imposes strict restrictions on the ROPs to be

estimated. Therefore, for future work, 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.

Zhang, Y., Huang, X., Hu, X., Wan, F., Lin, L., 2011. Direct relative orientation with four independent constraints. *ISPRS J. Photogramm. Remote Sens.* 66, 809–817.

ACKNOWLEDGEMENT

The information, data, or work presented herein was funded in part by the Advanced Research Projects Agency-Energy (ARPA-E), U.S. Department of Energy, under Award Number DE-AR0000593. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

REFERENCES

Faugeras, O.D., Maybank, S., 1990. Motion from point matches: multiplicity of solutions. *Int. J. Comput. Vis.* 4, 225–246.

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

Habib, A., Kelley, D., 2001. Automatic relative orientation of large scale imagery over urban areas using modified iterated Hough transform. *ISPRS J. Photogramm. Remote Sens.* 56, 29–41.

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

Heipke, C., 1997. Automation of interior, relative, and absolute orientation. *ISPRS J. Photogramm. Remote Sens.* 52, 1–19.

Horn, B.K., 1990. Recovering baseline and orientation from essential matrix. *J Opt Soc Am* 110.

Longuet-Higgins, H.C., 1987. A computer algorithm for reconstructing a scene from two projections. *Read. Comput. Vis. Issues Probl. Princ. Paradig. MA Fisch. O Firschein Eds* 61–62.

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

Mikhail, E.M., Bethel, J.S., McGlone, J.C., 2001. *Introduction to modern photogrammetry*. John Wiley & Sons Inc.

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

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

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

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