

Linear Approach for Initial Recovery of the Exterior Orientation Parameters of Randomly Captured Images by Low-Cost Mobile Mapping Systems

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

^a Lyles School of Civil Engineering, Purdue University,
47906, West Lafayette, Indiana, USA
- (he270, ahabib)@purdue.edu

Commission I, WG I/3

KEY WORDS: Exterior Orientation, Relative Orientation, Linear Approaches, Bundle Adjustment

ABSTRACT:

In this paper, we present a novel linear approach for the initial recovery of the exterior orientation parameters (EOPs) of images. Similar to the conventional Structure from Motion (SfM) algorithm, the proposed approach is based on a two-step strategy. In the first step, the relative orientation of all possible image stereo-pairs is estimated. In the second step, a local coordinate frame is established, and an incremental image augmentation process is implemented to reference all the remaining images into a local coordinate frame. Since our approach is based on a linear solution for both the relative orientation estimation as well as the initial recovery of the image EOPs, it does not require any initial approximation for the optimization process. Another advantage of our approach is that it does not require any prior knowledge regarding the sequence of the image collection procedure, therefore, it can handle a set of randomly collected images in the absence of GNSS/INS information. In order to illustrate the feasibility of our approach, several experimental tests are conducted on real datasets captured in either a block or linear trajectory configuration. The results demonstrate that the initial image EOPs obtained are accurate and can serve as a good initialization for an additional bundle adjustment process.

1. INTRODUCTION

Nowadays, 3D modelling of objects can be achieved using either passive or active remote sensing systems. Active sensors, such as laser scanners, are able to directly provide precise and reliable 3D information of scanned objects. However, the derived point cloud usually lacks spectral information (especially when dealing with collected data by mobile platforms). On the other hand, passive sensors, which commonly use digital frame cameras, can be incorporated for 3D reconstruction while providing spectral information of the derived coordinates. Compared to active sensors, the spectral information from passive sensors would allow for the derivation of better and more reliable semantic information pertaining to the reconstructed objects (e.g., the type and condition of mapped objects could be easily derived from the resulting 3D models). Therefore, passive-sensor-based reconstruction still remains the most complete, economical, flexible, and widely-used 3D modelling option in many areas (Remondino and El-Hakim, 2006).

3D reconstruction from digital images captured by passive sensors requires the knowledge of the Interior Orientation Parameters (IOP) of the utilized camera, the Exterior Orientation Parameters (EOP) of the involved images, and the corresponding points/features in the set of overlapping images. The IOP of the utilized camera can be derived from a camera calibration process. The EOP of the involved imagery can be either derived through an indirect geo-referencing procedure using tie and control points or a direct geo-referencing process through the implementation of a GNSS/INS unit on-board the mapping platform. While the latter approach provides practical convenience in terms of simplifying the geo-referencing process, it requires significant initial investment for the acquisition of the high-end GNSS/INS Position and Orientation System (POS) – especially, when seeking high level of reconstruction accuracy. Therefore,

significant research efforts have been exerted towards the development of automated procedures for 3D reconstruction and derivation of the geo-referencing parameters of the involved imagery in the absence of a GNSS/INS unit or in the presence of less accurate POS information from consumer-grade GNSS/MEMS units.

As far as the 3D reconstruction is concerned, one of two approaches could be adopted. In the first approach, a two-step procedure is adopted for the 3D reconstruction assuming the availability of prior knowledge regarding the EOP of the involved images. In the first step, corresponding features are identified in the set of available images (i.e., the matching problem). In the second step, the 3D positions of the matched points are derived using a simple intersection procedure that incorporates their image coordinates, IOP of the utilized camera, and the EOP of the involved images (Kraus, 2007). The main drawback of this procedure is the reliance on the availability of highly accurate EOP, which could be only available through the utilization of high-end POS units. Within the second approach for 3D reconstruction, which was mainly initiated by the computer vision research community, the feature matching and EOP recovery are simultaneously established. A commonly used procedure in this approach is known as Structure from Motion (SfM), which is usually based on a 3-step strategy. In the first step, the relative orientation parameters relating stereo-images or image triplets are initially estimated using automatically identified point and/or line features. Then, in the second step, a reference coordinate system is established and utilized to define the position and orientation parameters for the involved imagery using the derived relative orientation parameters in the first step as well as 3D coordinates of the matched points. Finally, in the third step, a bundle adjustment procedure is usually implemented to refine the derived information in the second step (Triggs et al.,

* Corresponding author. This is useful to know for communication with the appropriate person in cases with more than one author.

2000). Compared to the first approach for 3D reconstruction, SfM is more advantageous since it allows for 3D mapping in the presence or absence of GNSS/INS units.

In order to overcome the limitations of using GNSS/INS units, we adopt a Structure-from-Motion-based procedure for 3D reconstruction in this paper. A novel linear approach, which aims at the recovery of the EOPs of randomly captured images, is proposed. This proposed approach is tested on real image datasets, which were acquired by a calibrated camera – i.e., a camera with known IOPs. The remainder of the paper presents the proposed approach in more details. First, a literature review of related work is given. Then, the proposed methodology is introduced. Finally, experimental results and conclusions are presented for future work.

2. RELATED WORK

SfM 3D reconstruction approaches usually use either a sequential or hierarchical approach to initially estimate the EOP of the involved images – the focus of the second step of the aforementioned 3-step strategy – in an incremental manner. For example, Snavely et al. (2006) proposed an incremental SfM procedure, in which images are added one by one into the reference frame. In their method, the reference frame is established from a single pair of images that has a large number of matched points/features and a long baseline. Then, a new image is incrementally augmented to the reference frame. The EOPs of the augmented image are estimated using the reconstructed 3D points from the first pair through a Direct Linear Transformation (DLT) procedure. Fitzgibbon and Zisserman (1998) developed a hierarchical approach to recover the EOPs for either closed or open set of acquired images. In this method, trifocal tensors are estimated for all consecutive image triplets. Then, a hierarchical approach is applied to gradually integrate the image triplets to subsets. Finally, these subsets are augmented into a single block. For either the sequential or hierarchical approaches, intermediate bundle adjustment – which is time consuming – is implemented to ensure successful augmentation of the individual images into the final image block. Other than the expensive intermediate bundle adjustment, current incremental SfM methods suffer from severe error propagation during the sequential image augmentation process. Therefore, some global methods have been developed to simultaneously establish the EOPs for all the available images.

Martinec and Pajdla (2007) developed a two-step global method to solve for the EOPs of the images. First, they determine the rotational component of the EOPs of the images. Then, the spatial component of the EOPs is estimated in the second step. However, this two-step approach cannot deal with images that have been captured from an almost linear trajectory – those images will be denoted here forth as linear-trajectory images. Jiang et al. (2013) proposed another two-step global linear method for the estimation of EOPs of available images. To handle linear-trajectory images, they use the baseline/depth ratios from neighbouring stereo-pairs for the image augmentation process. However, the estimated baseline/depth ratios may not be accurate.

3. METHODOLOGY

Image-based 3D reconstruction requires the availability of accurate image EOPs. In both photogrammetric and computer vision research communities, bundle adjustment process, which simultaneously refines the 3D coordinates of reconstructed object points as well as the EOPs of images, is usually applied as the final step of image-based 3D reconstruction. However, bundle

adjustment is a nonlinear optimization algorithm, and it requires close initial approximation of the inputs. In this paper, the proposed approach is based on a linear solution for the initial recovery of the image EOPs. This approach includes two steps. In the first step, the relative orientation parameters relating stereo-images are derived from the essential matrix that is directly computed from conjugate point features. In the second step, a local coordinate frame is initially established, and then the EOPs of remaining involved images are sequentially recovered in an incremental augmentation process. A workflow of the proposed approach is shown in Figure 1.

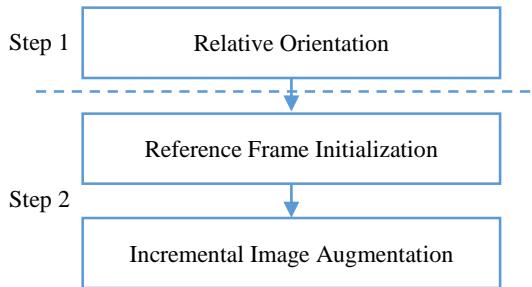


Figure 1. The proposed workflow for initial recovery of the image EOPs

3.1 Relative Orientation

In this section, the proposed relative orientation parameters (ROPs) estimation is presented.

3.1.1 SIFT Matching:

The estimation of ROPs requires the identification of conjugate points and/or line features in the set of available images. In this research, the Scale-Invariant Feature Transform (SIFT) features (Lowe, 1999), which are invariant to image scaling and rotation, are used to identify corresponding point features among the available stereo-images. The proposed SIFT feature matching process is implemented in two steps.

At the first step, the SIFT operator is applied for the identification of corresponding points in the available stereo-pairs. Initial SIFT feature correspondences are determined through a Euclidean-distance-based nearest neighbour matching approach, in which the Euclidean distances between the descriptors of the SIFT features are computed. At the second step, incorrect feature correspondences are detected through forward/backward consistency check. As shown in Figure 2, in the forward consistency check process, P1 matches Q1, and P2 matches Q2; while in the backward consistency check process, Q1 matches P1, and Q2 matches P3. Finally, (P1, Q1) is accepted as a correct conjugate pair, while (P2, Q2) and (Q2, P3) are discarded as mismatches.

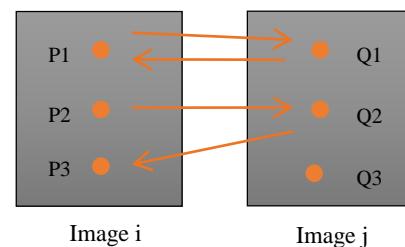


Figure 2. Forward/backward consistency check for an image pair

3.1.2 Initial Estimation of ROPs from the Essential Matrix:

The essential matrix is a 3-by-3 matrix comprised from the ROPs relating two stereo-images. More specifically, for a calibrated camera (i.e., a camera with known IOPs), if p_l and p_r are image coordinates of the identified conjugate points in the two stereo-images, the mathematical model relating the essential matrix and the conjugate points can be expressed in Equation 1.

$$p_l E p_r = 0 \quad (1)$$

The essential matrix can be computed from a set of conjugate points using either the 8-point or 5-point algorithm (Nistér, 2005). Meanwhile, the RANSAC algorithm (Fischler and Bolles, 1981) is usually implemented with the 8-point or 5-point algorithm for identifying incorrect feature correspondences.

Once the essential matrix is established, the initial rotation matrix R and translation vector T can be derived through a Singular Value Decomposition (SVD) process that is implemented according to the following sequence:

1. Apply an SVD on the essential matrix E :

$$E = U \sum V^T \quad (2)$$

Where U and V are two 3-by-3 orthogonal matrices, and \sum is a 3-by-3 diagonal matrix. According to the characteristics of the essential matrix, the diagonal entries of \sum are two identical and one zero values.

2. Define two matrices W and Z :

$$W = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } Z = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (3)$$

3. Compute two possible solutions for the rotation matrix R :

$$R = UWV^T \text{ and } R = UW^TV^T \quad (4)$$

4. Compute two possible solutions for the translation vector T :

$$T = VZV^T \text{ and } T = -VZV^T \quad (5)$$

As shown in Figure 3, four possible solutions for the rotation matrix R and translation vector T can be derived from the SVD method. Then, two additional constraints can be enforced to identify the one valid solution out of the four:

1. The two light rays connecting the object point and perspective centres should be on the same side of the baseline.
2. The object points should have depth consistent with the definition of the coordinate system.

Instead of using the SVD method, Horn (1990) proposed another approach for recovering the rotation and translation parameters from the essential matrix. In this method, the rotation matrix and translation vector are separately estimated. Similarly, four possible solutions are derived from this approach.

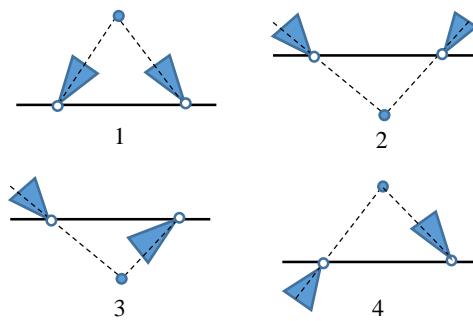


Figure 3. Four possible solutions for rotation matrix R and translation vector T derived from the essential matrix E

3.1.3 Refinement by Co-planarity Model:

The rotation matrix R and translation vector T derived from the essential matrix give a good initial approximation for the ROPs of the stereo-pairs. Then, the well-known co-planarity model (e.g., Mikhail et al., 2001) is adopted to provide accurate solution of the ROPs. The co-planarity model is presented by Equation 6 for a dependent relative orientation.

$$p_l^T \begin{bmatrix} 0 & Tz & -Ty \\ -Tz & 0 & Ty \\ Ty & -Tx & 0 \end{bmatrix} R_r^l p_r = 0 \quad (6)$$

In this equation, $p = (x, y, -c)^T$ represents the image coordinates corrected for principal offset and lens distortions. The rotation matrix R_r^l describes the rotation relating the right image to the left one. (Tx, Ty, Tz) are the translation vector connecting the two images of the stereo-pair in question. Based on the direction of the baseline, either Tx or Ty is assigned an arbitrary value since the co-planarity model does not recover the scale. The co-planarity model can be solved through a least-squares adjustment. At least five pairs of corresponding points are needed to recover the ROPs for a stereo-pair (i.e., three rotation angles and two translations).

3.2 Local Frame Initialization

Once the ROPs of all possible stereo-pairs are estimated, an incremental approach is proposed for the initial recovery of the image EOPs. This incremental approach is initiated by defining a local coordinate frame. Then, all the images are sequentially augmented into a final image block or trajectory.

In this paper, the local coordinate system is established using an image triplet that satisfies both a large number of corresponding points and good compatibility configuration. The first condition can be easily satisfied by maximizing the total number of corresponding points within the image triplet. The second condition is evaluated through a method referred to as compatibility analysis, which was developed by He and Habib, in 2014.

The compatibility analysis assumes a triangular relationship within an image triplet. Then, by using the estimated ROPs, both the rotational and translational errors within the image triplet are computed. These two types of errors indicate the image compatibility within the image triplet. In practice, an image triplet with good compatibility configuration usually leads to small rotational and translation errors. Therefore, the compatibility analysis of the initial image triplet can be performed by selecting the image triplet with the minimum rotational and translation errors.

Since the proposed compatibility analysis assumes a triangular relationship among the three images, it cannot handle a set of images along a linear trajectory configuration. To resolve this problem, rather than selecting an initial image triplet, a single pair of images that satisfies a large number of corresponding points as well as good intersection geometry is utilized to establish the local coordinate frame. In this case, the initial stereo-pair is evaluated through the baseline/depth ratio, which is derived from the estimated stereo-pair ROPs and 3D coordinates of reconstructed object points.

3.3 Incremental Image Augmentation

Once the local coordinate frame is established, the remaining images are sequentially augmented into the final image block or trajectory. The proposed approach utilizes a geometric structure, which is referred to as the reverse tree structure for the augmented images (Martinec and Pajdla, 2007), as the basis for the EOP recovery. In this paper, the proposed reverse tree-structure-based approach incorporates a linear approach for the estimation of the image EOPs. In the proposed procedure, the rotational and positional components of the image EOPs is conducted separately according to the following sequence.

3.3.1 Rotational Component Estimation:

The proposed approach for rotational component estimation is similar to the one suggested by Martinec and Pajdla, 2007. This approach is based on the relative orientation between the images of a stereo-pair. For a given stereo-pair including a referenced image i and an unreferenced image j , the relative orientation between these images and their orientation relative to the local coordinate frame can be formulated as Equation 7.

$$(R_i^l)^T = R_j^l (R_i^l)^T \quad (7)$$

Where R_j^l represents rotation matrix relating images i and j , which can be derived from the relative orientation estimation as explained in Section 3.1. R_i^l represents rotation of the referenced images i in the local coordinate frame. From a single stereo-pair, a system of nine linear equations can be established. Then, the nine unknown elements of the rotation matrix R_j^l , which represents the rotation of the unreference image j in the local coordinate frame, can be estimated through a least-squares adjustment. In this research, all referenced images which have overlap with the unreference image j is utilized to estimate the unknown rotational component. However, the orthogonal constraint of the rotation matrix is not enforced in Equation 7.

3.3.2 Positional Component Estimation:

Once the rotational component estimation is completed, the positional components of the image EOPs can be estimated. Based on the proposed reverse tree structure, the positional component of the image EOPs can be derived through an intersection of multiple vectors, which are the translation vectors connecting the referenced and unreference images. Meanwhile, in order to reduce the effects of error propagation, the proposed approach is based on augmenting images that exhibit the best compatibility with previously referenced images during the incremental image augmentation. Similar to the approach introduced in Section 3.2, the rotational and translational errors that are derived from the estimated image EOPs are utilized to evaluate the image compatibility configuration. In this research, at each step of the image incremental augmentation, only the

image that exhibits the highest compatibility with the previously referenced imagery is selected and referenced into the local frame.

3.4 Linear Trajectory Configuration

It is important to note that the utilized multi-vector intersection model in Section 3.3.2 assumes non-collinear relationship within the reverse tree structure. Given a set of images captured in a linear trajectory configuration, the proposed approach for positional component estimation suffers from a rank-deficiency, and cannot recover the position of the involved images.

In order to recover the positions of the linear-trajectory images, conjugate points, which are obtained from the aforementioned SIFT matching process, are first tracked through all images in the reverse tree structure. Then, the 3D object coordinates of these conjugate points are derived through a linear spatial intersection. Finally, the position of the unreferenced linear-trajectory image is recovered from these reconstructed object points. In the proposed approach, at least two object points are required for the positional component estimation. In order to obtain more reliable estimation, more object points with good spatial distribution are needed. Different from the single photo resection approach, which simultaneously estimates the rotational and positional components of the image EOPs, we only recover the positional components for the linear-trajectory image EOPs, while the rotational component of the image EOPs can be derived through the approach introduced in Section 3.3.1.

4. EXPERIMENTAL RESULTS

To illustrate the feasibility of the proposed procedure, we conducted several tests on real image datasets that have been captured in different configurations.

4.1 Dataset Description

The test site involved in the experimental tests was a building with complex roof structure. Two sets of images were captured at the test site. The first image dataset includes a block of 28 images captured by a low-cost DJI Phantom 2 UAV with a GoPro 3 camera (see Figure 7). The second image dataset consists of 21 images captured by a hand-held Canon Rebel T3 digital camera. For both datasets, the utilized cameras are calibrated (i.e., the IOPs of the GoPro camera and the Canon Rebel T3 camera are known). Figure 8 illustrates the sample images captured by the UAV and hand-held canon cameras.



Figure 4. The DJI Phantom 2 UAV drone with GoPro camera



Figure 5. (a) A sample UAV image and (b) a sample image captured by the hand-held Canon Rebel T3 digital camera

4.2 Results of the Proposed Approach

The proposed procedure for the initial recovery of the EOPs was tested on both image datasets. Figure 9 illustrates the estimated image position and orientation as well as the reconstructed sparse point cloud from the UAV image dataset. Figure 10 illustrates the results obtained from the Canon image dataset.

For both image datasets, we apply a final bundle adjustment using the initial EOPs that are derived through the proposed approach. The re-projection errors of both image datasets obtained from the bundle adjustment process are less than half a pixel. This quality indicates that the initial EOPs recovered from the proposed procedure can serve as a good initialization for a final bundle adjustment. Since the two available image datasets for the experiments are based on different image configurations, the results also indicate that the proposed procedure can handle sets of images that are captured with either a block or linear trajectory configuration.

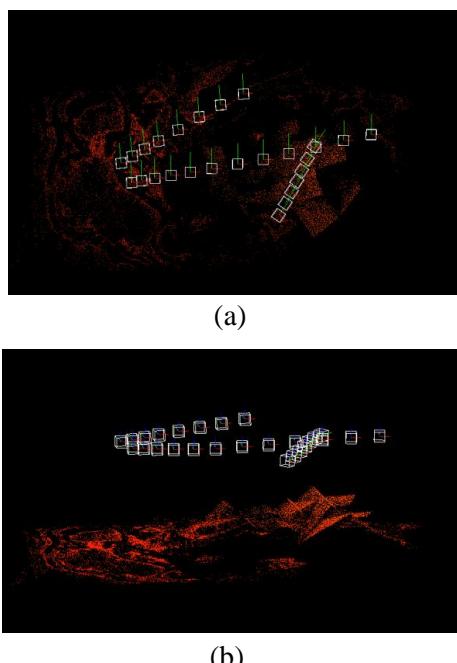


Figure 6. A top-view and (b) a side-view of the UAV image dataset and the reconstructed sparse point cloud

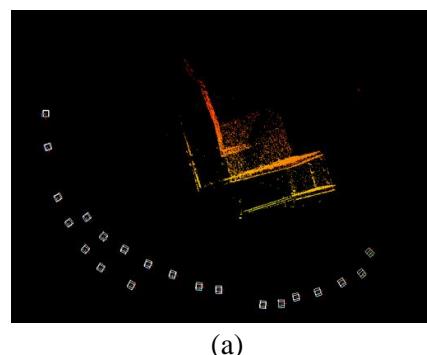


Figure 7. (a) A top-view and (b) a side-view of the Canon image dataset and the reconstructed sparse point cloud

5. CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK

This paper presents a novel approach for the initial recovery of image EOPs. The experimental results demonstrate that the proposed approach has the following characteristics:

1. It is based on a linear approach for the estimation of the ROPs among the available stereo-pairs as well as the sequential estimation of the rotational and positional components of the image EOPs relative to a local coordinate frame,
2. The incremental image augmentation is based on augmenting images that exhibit the highest compatibility with the previously referenced imagery. It reduces the effects of error propagation during the recovery of image EOPs.
3. It can handle a set of randomly collected image set (i.e., it doesn't assume any prior knowledge regarding the sequence of the image collection procedure). This would be advantageous in dealing with images that have been captured from different campaigns,
4. It can handle images that have been captured in a block or linear trajectory configuration.

It is important to notice that the proposed approach follows an incremental approach for image augmentation. Therefore, it may be less efficient compared to global approaches, especially for datasets containing a large number of images. For future work, more experiments will be carried out to compare obtained results from our approach with results derived from other approaches. Meanwhile, to improve the efficiency of this proposed approach, we also plan to investigate parallel processing and GPU, which can efficiently reduce the computational time of the proposed approach.

ACKNOWLEDGEMENT

The authors would like to thank Abdullah Rawabdeh for providing the UAV system and collecting the experimental data. The authors would like also to thank Tecterra for the financial support of this research work.

REFERENCES

- Fitzgibbon, A.W., Zisserman, A., 1998. Automatic camera recovery for closed or open image sequences, in: Computer Vision—ECCV'98. Springer, pp. 311–326.
- He, F., Habib, A., 2014. Automatic orientation estimation of multiple images with respect to laser data. Presented at the ASPRS 2014 Annual Conference, Louisville, USA.
- Horn, B.K.P., 1990. Recovering baseline and orientation from essential matrix. *J. Optical Society of America*.
- Jiang, N., Cui, Z., Tan, P., 2013. A Global Linear Method for Camera Pose Registration, in: Computer Vision (ICCV), 2013 IEEE International Conference on. IEEE, pp. 481–488.
- Kraus, K., 2007. Photogrammetry: geometry from images and laser scans. Walter de Gruyter.
- Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* 60, 91–110.
- Martinec, D., Pajdla, T., 2007. Robust rotation and translation estimation in multiview reconstruction, in: Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on. IEEE, pp. 1–8.
- 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.
- Remondino, F., El-Hakim, S., 2006. Image-based 3D Modelling: A Review. *Photogramm. Rec.* 21, 269–291.
- Snavely, N., Seitz, S.M., Szeliski, R., 2006. Photo tourism: exploring photo collections in 3D. *ACM Trans. Graph. TOG* 25, 835–846.
- Triggs, B., McLauchlan, P.F., Hartley, R.I., Fitzgibbon, A.W., 2000. Bundle adjustment—a modern synthesis, in: Vision Algorithms: Theory and Practice. Springer, pp. 298–372.