



Article

A Quadrifocal Tensor SFM Photogrammetry Positioning and Calibration Technique for HOFS Aerial Sensors

Tao Wang , Yan Zhang ^{*}, Yongsheng Zhang, Ying Yu, Lei Li, Shaocong Liu, Xiang Zhao, Zhenchao Zhang and Longhui Wang

Geospatial Information Institute, Information Engineering University, Zhengzhou 450001, China; wangtaoyln@163.com (T.W.); yszhang2001@vip.163.com (Y.Z.); yuying5559104@163.com (Y.Y.); 31110100798@zju.edu.cn (L.L.); scholar_liusc@163.com (S.L.); zx15538019261@163.com (X.Z.); zhzhc_1@163.com (Z.Z.); wlhonline@163.com (L.W.)

* Correspondence: zhangyanxz7806@163.com; Tel.: +86-186-9580-9656

Abstract: Nowadays, the integration between photogrammetry and structure from motion (SFM) has become much closer, and many attempts have been made to combine the two approaches to realize the positioning, calibration, and 3D reconstruction of a large number of images. For the positioning and calibration of high oblique frame sweep (HOFS) aerial cameras, a quadrifocal tensor SFM photogrammetry technique is proposed to resolve the positioning and calibration task of such cameras. It adopts the quadrifocal tensor idea into the OpenMVG SFM pipeline to solve the complexity problem caused by the small single-viewing imaging area and the high image overlapping ratio. It also integrates the photogrammetry iteration idea into the OpenMVG SFM pipeline to enhance the positioning and calibration accuracy, which includes a coarse to fine three-stage Bundle Adjustment (BA) processing approach. In this paper, the overall workflow of the proposed technique was first introduced in detail, from feature extraction and image matching, relative rotation and translation estimation, global rotation and translation estimation, and the quadrifocal tensor model construction to the three-stage BA process and calibration. Then, experiments were carried out in the Zhengzhou area, implementing four types of adjustment methods. The results suggest that the proposed quadrifocal tensor SFM photogrammetry is suitable for large tilt frame sweep camera positioning and calibration without prior information on detailed camera intrinsic parameters and structure. The modifications made to the OpenMVG SFM pipeline enhanced the precision of image positioning and calibration and provided the precision level of professional photogrammetry software.

Keywords: a quadrifocal tensor; SFM photogrammetry; oblique photogrammetry; frame sweep aerial sensor; VisionMap A3 edge sensor



Citation: Wang, T.; Zhang, Y.; Zhang, Y.; Yu, Y.; Li, L.; Liu, S.; Zhao, X.; Zhang, Z.; Wang, L. A Quadrifocal Tensor SFM Photogrammetry Positioning and Calibration Technique for HOFS Aerial Sensors. *Remote Sens.* **2022**, *14*, 3521. <https://doi.org/10.3390/rs14153521>

Academic Editors: Daniel Spengler and Mohammadmehdi Saberion

Received: 1 June 2022

Accepted: 20 July 2022

Published: 22 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Given its high spatial resolution and great flexibility, digital frame cameras are widely used in aerial photogrammetry and remote sensing for topographic mapping, archaeological discoveries, forestry and agricultural assessment, natural disaster monitoring, and ecosystem restoration [1–6]. To obtain broader ground coverage, frame digital cameras usually adopt the following technologies: super-wide frame cameras (e.g., Z/I DMC system [4,7]), multi-splicing frame cameras (e.g., Microsoft Ultracam system [4,8], Leica Citymapper system [9], and SWDC-4 cameras [10]), and frame sweep cameras. The representative aerial frame sweep cameras include the EOLOROP [11], DB-110 [12], and VisionMap A3 [13–16].

Most of these cameras are equipped with Position Orientation System (POS) equipment (including Global Navigation Satellite Systems (GNSS) and Inertial Measurement Units (IMU)) to provide initial camera poses and orientation. Due to significant imaging principles and camera inner structures, the specific camera imagery is always processed

through corresponding professional software, such as Ultramap software for Microsoft Ultracam cameras, HxMap software for Leica Citymapper cameras, and LightSpeed for VisionMap A3 cameras [16].

For super-wide and multi-splicing frame cameras, the geometric imaging relationship is relatively simple in nearly vertical photography. Even though the Leica Citymapper camera adopts an oblique photography approach, the angular relationship between different views is fixed. Thus, an accurate geometric imaging model can be established, and high-precision open-source positioning and calibration software can be developed, breaking the limitations of professional software.

In contrast, the frame sweep camera swings along the vertical direction of the flight swiftly for imaging, such as VisionMap A3. It achieves true high-tilt photogrammetry with a long focal length and an ultra-wide sweep angle up to 104 degrees. It can obtain a wide range of clear images at high altitudes, which is particularly useful for surveying, monitoring, and reconnaissance. The obtained multi-angle images are excellent for 3D reconstruction compared to high-resolution satellite images. However, the camera tilt angle varies considerably in dynamic imaging, causing complex geometric imaging relationships and inconsistent imaging scales within one sweep cycle. Moreover, the VisionMap A3 camera is not equipped with an IMU installment, providing no camera orientation information. These increase the difficulty in image positioning processing.

To realize the geometric processing for the VisionMap A3 camera in a more general and simplified way, the research project was implemented, and the performance of the proposed technique was comprehensively analyzed. The rest of the paper is organized as follows. The related works are discussed in the next section, and the VisionMap A3 system is introduced in Section 3. Section 4 presents the methods, including the ideas and workflow. The experimental results and the discussion are provided in Section 5, and a brief conclusion is presented in the last section.

2. Related Work

Photogrammetric techniques have been used to achieve open-source positioning and calibration of the VisionMap A3 camera, but the results have largely been unstable. The main goal of photogrammetry is to generate accurate camera poses (i.e., positions and orientations), intrinsic camera parameters, and 3D points. The processing approach includes matching, aero-triangulation (also called bundle adjustment), orthoimage generation, and digital elevation model (DEM) generation [1–5,17]. Complex geometric imaging relationships and inconsistent imaging scales make it difficult to set up accurate geometric imaging models; thus, many problems occur in processing A3 images by the photogrammetry scheme.

Aside from photogrammetry, SFM has become a common approach in aerial imagery processing. Photogrammetry and SFM approaches have many similarities in feature and bundle adjustment. SFM also computes camera poses, intrinsic parameters, and 3D points cloud [18–20]. SFM methods can be categorized into five types: global SFM, incremental SFM, hierarchical SFM, hybrid SFM, and semantic SFM. Global SFM processes all images simultaneously, first solving global camera rotation using the rotation consistency, then calculating camera displacement, and finally applying BA optimization. While the global strategy significantly improves processing efficiency, it is extremely sensitive to external points. Its performance is unstable in many applications, and its reconstruction accuracy is largely unsatisfactory [21–28].

Incremental SFM starts with two or three view reconstructions, gradually adds new views, and then applies BA operation after each addition. This processing method is robust with high reconstruction accuracy, and most popular SFM pipelines employ incremental approaches. However, incremental SFM has drift risk due to the accumulation of errors [29–37]. Hierarchical SFM is the revision of the traditional incremental SFM. It divides the large-scale datasets into N interrelated sub-datasets, and then parallel incremental SFM processing is performed on the sub-datasets. Finally, the sub-datasets are

merged [38–41]. Hybrid SFM combines the advantages of global SFM and incremental SFM. It solves camera rotation using the global approach and then calculates camera displacement incrementally [42–44]. Semantic SFM uses the semantic label to detect the corresponding models and facilitate image matching and camera calibration. Scene segmentation can be realized in SFM reconstruction [45–47].

The well-known SFM pipelines include Bundler [31,32], VisualSFM [35], COLMAP [34,37], and OpenMVG [29,36]. The properties of these pipelines are summarized in Table 1.

Table 1. Popular SFM Pipelines Analysis.

Pipeline	Properties
Bundler	An early incremental SFM pipeline originated from Photo Tourism Project. The time complexity is $O(n^4)$, the output is sparse point clouds, and the operation is complex.
VisualSFM	A revision on Bundler and the time complexity reaches $O(n^1)$, the output is dense point clouds with higher accuracy, and the operation is much simpler.
COLMAP	COLMAP improves scene graph verification, next best view selection, sampling-based triangulation, and bundle adjustment; it performs better in terms of completeness and robustness. Its efficiency outperforms OpenMVG but does not support POS data input.
OpenMVG	OpenMVG adds an efficient contrario trifocal tensor estimation method and translation registration technique into the original SFM pipeline. It outperforms many incremental and global pipelines in terms of accuracy and running times. The reprojection error and the drift are smaller than COLMAP, and it supports POS data input.

Notes: n is the number of images.

At present, the integration between photogrammetry and SFM has become more pronounced. SFM photogrammetry combines the two approaches to realize the positioning, calibration, and 3D reconstruction of a large number of ordered and disordered images [48].

Photogrammetry often relies on initial POS data and intrinsic camera parameters and can reach higher accuracy through iterative least-squares adjustment, condition adjustment, or indirect adjustment. SFM takes homogeneous coordinates in expression and does not require specific initial values of POS data and intrinsic camera parameters. In image pose recovery and 3D reconstruction, SFM uses the R matrix to model the attitude angle system, which does not involve selecting a specific angle rotation system; thus, it can overcome challenges caused by the swift change of the sweep angle in the ultra-wide imaging range. It also avoids conversion among different coordinate systems, resulting in much simpler operations and faster convergence. Thus, the SFM pipeline was chosen to solve the positioning and calibration task of the VisionMap A3 camera.

By analyzing Table 1 it can be determined that the OpenMVG pipeline has smaller reprojection and drift errors and supports POS data input, which is especially advantageous for remote sensing imagery. Thus, the OpenMVG pipeline was chosen for the geometric processing of VisionMap A3 images, addressing the limitations of the professional LightSpeed software.

Wu et al. [35] introduced a novel BA strategy that provides a good balance between SFM reconstruction speed and accuracy and maintains high accuracy by regularly retriangulating feature matches that fail to triangulate. Wu's feature detection and description algorithm was adopted into the OpenMVG pipeline. Moulon et al. [36] proposed a new global calibration approach based on the fusion of relative motions between image pairs, presenting an efficient contrario trifocal tensor estimation method and translation registration technique for accurate camera positions recovery in OpenMVG. We followed the tensor cluster idea and designed the quadrifocal tensor-based BA method.

Two main improvements were introduced to the original OpenMVG pipeline. First, the VisionMap A3 frame sweep camera has a small single-viewing imaging area. The high image overlapping ratio and the high number of images enhance the processing complexity. To reduce the complexity and strengthen the robustness, the quadrifocal tensor was added to the OpenMVG pipeline, which deals with the images in the quadrifocal tensor

model cluster. The quadrifocal tensor model performs high oblique frame sweep camera positioning and calibration processing as an independent unit. Second, the photogrammetry iteration idea was introduced into the OpenMVG pipeline to enhance the positioning and calibration accuracy. A coarse to fine three-stage BA processing approach was proposed to deal with the pose parameters and the intrinsic parameters.

The quadrifocal tensor-based positioning and calibration method first carries out feature extraction and image matching. Relative rotation and translation estimation and global rotation and translation are performed before establishing the quadrifocal tensor model. The three-dimensional coordinates of matching points in each quadrifocal tensor model are calculated, and the BA cost function is finally set up with the quadrifocal tensor model as the independent unit. Real VisionMap A3 data is then used to evaluate whether the proposed approach can solve the high oblique frame sweep camera position and calibration problem in a more general and simplified way.

3. VisionMap A3 System Introduction

The VisionMap A3 system is a fully automated mapping system established in 2004. The system consists of an airborne digital step-framing double lens metric camera and a ground processing system. The airborne system consists of dual CCDs with two 300 mm lenses (Figures 1 and 2), a fast compression and storage unit, and a dual-frequency GPS. The long focal length yields comparatively high ground resolution when flying at high altitudes, enabling efficient photography of large areas in high resolution.



Figure 1. VisionMap A3.



Figure 2. Dual Lenses of VisionMap A3.

During the flight, a sequence of frames is exposed in a cross-track direction at a very high speed to provide a very wide angular coverage of the ground. The two lenses of the camera simultaneously sweep across the flight direction from one side to the opposite side, with each CCD capturing about 27 frames (54 frames for two CCDs) and having a

maximum sweep angle of 104 degrees. After completing the first sweep, the lenses return to the start position in preparation for the next sweep. The sweep back time is 0.5 s. Each CCD captures seven frames per second (one frame in 0.142 s); therefore, a single sweep is completed in approximately 3.6 s. The time between sweeps depends largely on the aircraft speed, flight altitude, and the required overlap between two consecutive sweeps. For more details on the technical parameters, refer to Pechatnikov et al. [13].

Two adjacent single frames along the flight form a double frame (Figures 3 and 4). The overlap between two adjacent single frames is about 2% (~100 pix), while the overlap between two adjacent double frames across the flight direction is about 15%. The overlap between two consecutive sweeps along the flight direction is typically 56%, but this value may vary depending on the defined specifications of the aerial survey. Between two consecutive flight lines, the overlap is generally 50% to enable stereo photogrammetric mapping. All overlaps are determined during flight planning and may be altered during the flight. VisionMap A3 camera provides orthogonal coverage of the nadir area and oblique coverage of the remainder of the sweep image. As all images participate in all stages of the analytical computations, after performing matching and block adjustment, accurate solutions for all images, including the oblique images, can be obtained. The generation of accurately solved oblique images simultaneously with regular verticals is a unique and highly important feature of the A3 system.

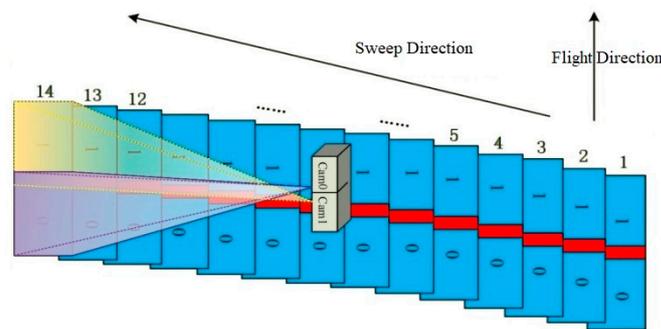


Figure 3. Flights with VisionMap A3. Number 0 represents images captured by lenses 0, number 1 represents images captured by lenses 1.

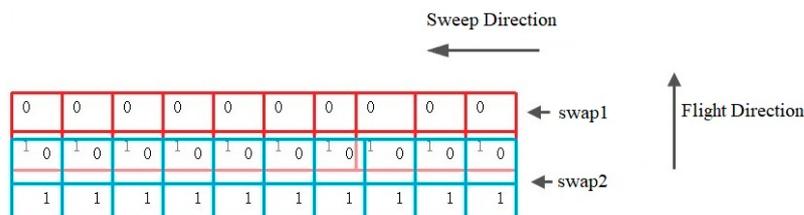


Figure 4. Relationship among Consecutive Swap. Number 0 represents images captured by lenses 0, number 1 represents images captured by lenses 1.

4. The Proposed Quadrifocal Tensor SFM Photogrammetry Positioning and Calibration Technique

The proposed quadrifocal tensor SFM photogrammetry positioning and calibration technique include the following steps. First, all aerial images are sorted by flight route according to imaging time, and the sorting relationship is established. The scale-invariant feature transform (SIFT) algorithm is then used to extract feature points on each view and build corresponding feature descriptors according to the image sorting relationship. The iterative feature matching process is then carried out under the Hamming distance criterion. The RANSAC algorithm is used to eliminate mismatches in the matching results, and the feature point correspondences on the pairwise overlapping images are determined. The pinhole imaging model is then established based on the unknown camera geometric distortion parameters, and the essential matrices between each pairwise overlapping

view are calculated using the successful matching point correspondences and redundant constraints. The relative spatial rotation and the translation relationship, i.e., the rotation matrix R and the translation vector t , are derived from the essential matrices.

After performing the Bayesian inferences, the initial global rotations are computed for each view in a global coordinate system, while the initial global translations are generated using the initial POS data or from the translation registration. The quadrifocal tensor model can then be established, and the initial three-dimensional coordinates of the matching point correspondences in each quadrifocal tensor model are calculated. Finally, the BA cost function based on the quadrifocal tensor model is constructed and used in the BA processing. The global rotation matrix and the translation vector for each view are updated, and the intrinsic parameters are calibrated simultaneously. The general workflow is listed in Figure 5.

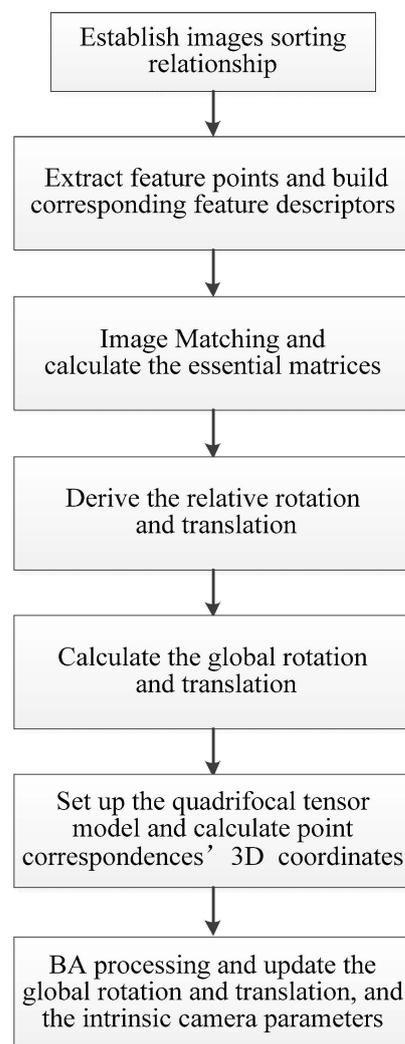


Figure 5. Workflow of the proposed quadrifocal tensor SFM photogrammetry positioning and calibration technique.

4.1. Feature Extraction and Image Matching

The SIFT algorithm is used to extract the feature points on each view and build corresponding feature descriptors simultaneously according to the image sorting relationship [49]. The Hamming distance is taken as the criterion for iterative image matching, and the RANSAC algorithm is adopted to eliminate a large number of mismatches in the matching results.

4.2. Relative Rotation and Translation Estimation

The pinhole camera model is used to describe the frame sweep aerial camera. The three-dimensional point $P(X, Y, Z)$ in the ground object space is projected onto the image plane and forms an image point $p(x, y, -f)$ through the pinhole projection. The geometric relationship between the object point and the image point can be described with the following model:

$$x = PX = K[R|t]X \quad (1)$$

where R is the rotation matrix, a 3×3 orientation matrix representing the direction of the camera coordinate system; t is the position vector; R and t are the camera exterior parameters; K is the camera intrinsic parameter matrix (also called the camera calibration matrix). The frame sweep camera positioning and calibration determine the R , t , and K matrices; the camera projection matrix is provided by $P = K[R|t]$.

The general form of the calibration matrix for a CCD camera is:

$$K = \begin{bmatrix} \alpha_x & & x_0 \\ & \alpha_y & y_0 \\ & & 1 \end{bmatrix}. \quad (2)$$

To increase generality, the skew parameter s can be added,

$$K = \begin{bmatrix} \alpha_x & s & x_0 \\ & \alpha_y & y_0 \\ & & 1 \end{bmatrix}. \quad (3)$$

Based on the successful matching point correspondences obtained in Section 4.1, the relative spatial rotation and translation between any two overlapping images can be calculated. The pairwise point correspondence p_1, p_2 meets the following epipolar constraint:

$$p_2^T K^{-T} E K^{-1} p_1 = 0 \quad (4)$$

In Equation (4), the points O_1, P , and O_2 are coplanar. E is the essential matrix, which is the outer product of t and R and is perpendicular to t and R . Both translation and rotation are included in the epipolar plane constraint, and the essential matrix E is calculated from the point correspondences. After obtaining the essential matrix E , singular value decomposition (SVD) is performed on the matrix E matrix to obtain R . Thus, the relative rotation and translation of the two overlapping images can be determined [18,29,36].

4.3. Global Rotation and Translation Estimation

Assuming that in the global coordinate system, the global rotation matrix of image i is R_i , that of image j is R_j , and the relative rotation is R_{ij} , which is obtained in Section 4.2. Three rotation matrices meet the consistency Equation (5), which is the basis for solving the global rotation.

$$R_j = R_{ij}R_i. \quad (5)$$

R_j is orthonormal, for $j = 1, \dots, m$.

While the RANSAC algorithm can remove most mismatches, there are remaining mismatches that cause deviations in the relative rotation estimation. As relative R_{ij} estimates may contain outliers, the global rotation estimation must be robust in identifying the global rotations and the inconsistent/outlier edges (false essential geometry). Moulon's method [36] is used in determining inconsistent relative rotations in Bayesian inference. The iterative use of the Bayesian inference, adjusted with the cycle length weighting, can remove most outliers, check all the triplets of the graph, and reject those with cycle deviations larger than 2° .

When the relative rotations are known, they form a tree graph with the views as vertices connected by an edge. Equation (5) can be solved by least-squares while satisfying

the orthonormality conditions, R can be expressed in quaternions [50], and Equation (5) is transformed into

$$r^j = r^{ij}r^i \tag{6}$$

where r^j and r^i are the unknown quaternions of the i th and j th view rotation, respectively, and r^{ij} is the known relative rotation between views i and j . Each quaternion can be considered a four-vector. Using known manipulations with quaternions, each equation in (6) can be rewritten as

$$\begin{pmatrix} r_0^j \\ r_x^j \\ r_y^j \\ r_z^j \end{pmatrix} = \begin{bmatrix} r_0 & -r_x & -r_y & -r_z \\ r_x & r_0 & -r_z & r_y \\ r_y & r_z & r_0 & -r_x \\ r_z & -r_y & r_x & r_0 \end{bmatrix} \begin{pmatrix} r_0^i \\ r_x^i \\ r_y^i \\ r_z^i \end{pmatrix} \tag{7}$$

where $r^i = r_0^i + ir_x^i + jr_y^i + kr_z^i$, with i, j , and k as imaginary units. There are $4m$ unknowns $r_0^1, r_x^1, r_y^1, r_z^1, \dots, r_0^m, r_x^m, r_y^m, r_z^m$ with constraints (7) for each view pair ij with a known rotation. After the solution, the quaternions can be easily made into units by dividing each by its Euclidean length.

After deciding on the global rotations, the next step is determining the global translations. Given a set of relative motion pairs (R_{ij}, t_{ij}) (rotations and translation directions solved in Section 4.2), the global location (T_1, \dots, T_m) of all views can be obtained. The different translation directions are reconciled in the global coordinates system.

$$\|T_j - R_{ij}T_i - \lambda_{ij}t_{ij}\| = 0, \quad \forall i, j \tag{8}$$

To solve m global translations and scale factors λ_{ij} , the solution for Equation (7) is optimized using the least-squares method or Moulon’s approach [36] under the l_∞ norm. For the VisionMap A3 edge camera, the GPS equipment provides the camera with position information in the world coordinate system, so the processed GPS information can be taken as initial values (T_1, \dots, T_m) .

4.4. Set up the Quadrifocal Tensor Model and Calculate Initial 3D Ground Point

Subsequently, the quadrifocal tensor model would be established, and the global translations and three-dimensional coordinates of the point correspondences would then be determined based on the quadrifocal tensor model. For the VisionMap camera, at a certain sweep angle, the cam0 and cam1 lenses obtain images I and I' , the corresponding projection planes are Π and Π' , and the projection centers are C and C' . The sweep angle changes swiftly. At the next imaging moment, the cam0 and cam1 lenses obtain images I'' and I''' , with the corresponding projection planes Π'' and Π''' and projection centers C'' and C''' . The straight line L in the object space is reflected on four image views (see Figure 6).

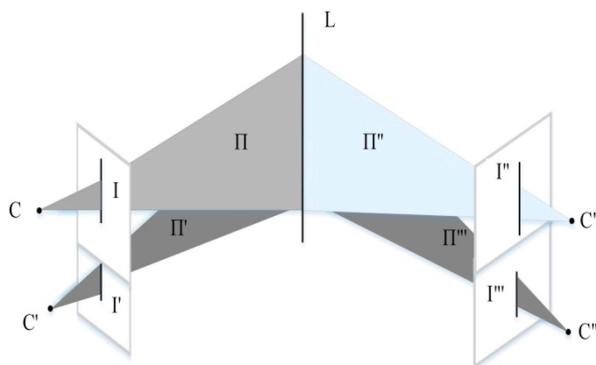


Figure 6. The Quadrifocal Tensor Model.

When the camera projection matrices of the four image views are A, B, C, and D, a three-dimensional ground point X projects onto the four image views, and there is a set of corresponding points $x \leftrightarrow x' \leftrightarrow x'' \leftrightarrow x'''$ across the four views. The four image views constitute a quadrifocal tensor model, and the projection relationship among the four views is provided by:

$$\begin{bmatrix} A & x \\ B & & x' \\ C & & & x'' \\ D & & & & x''' \end{bmatrix} \begin{bmatrix} X \\ -k \\ -k' \\ -k'' \\ -k''' \end{bmatrix} = 0 \tag{9}$$

where $k, k', k'',$ and k''' represent uncertain scale constants, and $a^i, b^i, c^i,$ and d^i are the row vectors of matrices A, B, C and D, respectively. The rank-4 tensor Q^{pqrs} is defined by

$$Q^{pqrs} = \det \begin{bmatrix} a^i \\ b^q \\ c^r \\ d^s \end{bmatrix}. \tag{10}$$

From Equation (9), the quadrilinear relationship can be derived in the following form for the quadrifocal tensor:

$$x^i x'^j x''^k x'''^l \varepsilon_{ipw} \varepsilon_{jqx} \varepsilon_{kry} \varepsilon_{lsz} Q^{pqrs} = 0_{wxyz} \tag{11}$$

where 0_{wxyz} is a zero tensor with four indices w, x, y and z ; $\varepsilon_{ipw}, \varepsilon_{jqx}, \varepsilon_{kry}$ and ε_{lsz} represent the product between the different vectors. Based on the global rotation and translation computed in Section 4.3, the camera projection matrices A, B, C, and D are determined, and the quadrifocal tensor model can then be established.

The next step is realizing 3D reconstruction for point correspondences in the quadrifocal tensor model. The ground point X projects across four views and forms the corresponding image points $x \leftrightarrow x' \leftrightarrow x'' \leftrightarrow x'''$. The geometric projection relationship between the ground point and the corresponding image point is listed as follows:

$$\begin{cases} x = AX \\ x' = BX \\ x'' = CX \\ x''' = DX \end{cases} \tag{12}$$

The plane coordinate of the image point x is (x, y). Taking the first expression in Equation (12) as an example, when the cross product is applied, and the homogeneous scalar factor is eliminated, the following equations can be obtained:

$$\begin{cases} x(A^{3T}X) - (A^{1T}X) = 0 \\ y(A^{3T}X) - (A^{2T}X) = 0 \\ x(A^{2T}X) - y(A^{1T}X) = 0 \end{cases} \tag{13}$$

Taking the first two equations, which are linearly independent, the four image points $x \leftrightarrow x' \leftrightarrow x'' \leftrightarrow x'''$ develop into eight equations

$$NX - L = 0N = \begin{bmatrix} xA^{3T} - A^{1T} \\ yA^{3T} - A^{2T} \\ x'B^{3T} - B^{1T} \\ y'B^{3T} - B^{2T} \\ x''C^{3T} - C^{1T} \\ y''C^{3T} - C^{2T} \\ x'''D^{3T} - D^{1T} \\ y'''D^{3T} - D^{2T} \end{bmatrix}. \tag{14}$$

Solving Equation (14) by the least-squares approach, the 3D coordinates for ground point X are generated.

4.5. Bundle Adjustment Based on the Quadrifocal Tensor Model

The BA processing is performed to update the rotation and translation and simultaneously calibrate the intrinsic matrix. The BA cost function based on the quadrifocal tensor model is defined in the form as

$$f(z) = \left\| \left(x_i - \frac{R_i^1 X + t_i^1}{R_i^3 X + t_i^3}, y_i - \frac{R_i^2 X + t_i^2}{R_i^3 X + t_i^3} \right) \right\| \tag{15}$$

where $(x_i, y_i), i \in (1, 2, 3, 4)$ is the observation value corresponding to the four image points $x \leftrightarrow x' \leftrightarrow x'' \leftrightarrow x'''$; $R_i, i \in (1, 2, 3, 4)$ is the corresponding rotation matrix to each view; $t_i, i \in (1, 2, 3, 4)$ is the corresponding translation vector; $\left(\frac{R_i^1 X + t_i^1}{R_i^3 X + t_i^3}, \frac{R_i^2 X + t_i^2}{R_i^3 X + t_i^3} \right)$ is the calculation value of the image point; z is the unknown parameter vector comprising the camera pose and calibration parameters (i.e., unknowns in rotation and translation) and the 3D ground point coordinates X.

The differences between the observed values and the calculation values of the image points are the reprojection error of the ground point, representing the errors contained in the rotation, translation, and camera calibration parameters and the ground point coordinates. When the cost function reaches the minimum value, the optimal solution is obtained for the pose parameters, calibration parameters, and 3D coordinates of ground points.

The BA process optimizes the following nonlinear least-squares BA cost function

$$z^* = \operatorname{argmin}_z \sum_{k=1}^n \|f_k(z)\|^2. \tag{16}$$

The BA cost function is optimized by the Levenberg–Marquardt (LM) approach. The LM algorithm decomposes the original nonlinear cost function (Equation (16)) into approximations of a series of regularized linear functions and $J(z)$ is the Jacobian matrix of $f(z)$. In each loop, the LM approach updates the linear least squares problem in the following form:

$$\delta^* = \operatorname{argmin}_\delta \|J(z)\delta + f(z)\|^2 + \lambda \|D(z)\delta\|. \tag{17}$$

If $\|f(z + \delta^*)\| < \|f(z)\|$, then the unknown parameter is updated as $z \rightarrow z + \delta^*$. $D(z)$ is the square root matrix of the matrix $J(z)^T J(z)$ and λ is the regularization parameter, which can be adjusted according to the approximation between $J(z)$ and $f(z)$. Solving Equation (17) is equivalent to solving the standard equation:

$$(J^T J + \lambda D^T D)\delta = -J^T f \tag{18}$$

where $(J^T J + \lambda D^T D)$ is the extended Hessian matrix. In the BA process, the unknown parameter z includes two parts: the camera pose and calibration parameters (z_c) and the 3D ground point coordinates (z_p). Similarly, D, δ, J can also be divided into two parts. Let

$U = J_c^T J_c$, $V = J_p^T J_p$, $U_\lambda = U + \lambda D_c^T D_c$, $V_\lambda = V + \lambda D_p^T D_p$, $W = J_c^T J_p$; Equation (18) can then be rewritten as the block structure linear system:

$$\begin{bmatrix} U_\lambda & W \\ W^T & V_\lambda \end{bmatrix} \begin{bmatrix} \delta_c \\ \delta_p \end{bmatrix} = - \begin{bmatrix} J_c^T f \\ J_p^T f \end{bmatrix}. \quad (19)$$

Applying the Gaussian elimination method to Equation (19), the 3D coordinates of the ground points can be eliminated, and a linear equation containing just the camera pose and calibration parameters can be obtained:

$$(U_\lambda - W V_\lambda^{-1} W^T) \delta_c = -J_c^T f + W V_\lambda^{-1} J_p^T f. \quad (20)$$

Therefore, δ_c can be determined by solving Equation (20) and δ_p can then be solved by reverse substitution.

$$\delta_p = -V_\lambda^{-1} (J_p^T f + W^T \delta_c). \quad (21)$$

The variables δ_c and δ_p are obtained using an iterative solution; after two loops, both can reach high precision. The quadrifocal tensor SFM positioning and calibration process for the frame sweep aerial camera is therefore completed.

To obtain better positioning and calibration precision, the OpenMVG SFM pipeline is improved according to the photogrammetry processing mode. In photogrammetric processing, position deviation is mostly caused by pose errors, while the intrinsic camera parameters generate minor systematic errors for both aerial and satellite-borne cameras [51,52]. So in the photogrammetry BA implementation, BA is always carried out in several stages. Pose parameters are dealt with first, while the intrinsic parameters are processed last. Here, three-stage BA processing is put forward. In the first stage, the intrinsic matrix K is set to E (identity matrix), the rotation R remains unchanged, and only the translation T and the 3D coordinates of ground points X are involved. In the second stage, the intrinsic matrix K remains as E , but the translation T , the rotation T , and the 3D coordinates of ground points X are involved in the BA process. In the last stage, all the unknowns, the intrinsic parameter K , the translation T , the rotation T , and the 3D coordinates of ground points X are adjusted.

5. Experiments and Analysis

5.1. Test Data

The VisionMap Edge A3 camera was used to photograph the Zhengzhou area in January 2020, with a total flight area of about 1000 square kilometers. Two flight missions were carried out at 2300 m flight altitude, generating a ground resolution of about 5 cm. Six routes (L_44, L_45, L_46, L_1, L_2, and L_3), located in the middle region and covering about 120 square kilometers, were used for the experiment (see Figure 7).

The upper half is selected for analysis, comprising 7928 images and covering a total area of 64 square kilometers. The study area (shown in Figure 8) includes plain regions and urban areas.



Figure 7. VisionMap Flight Coverage.

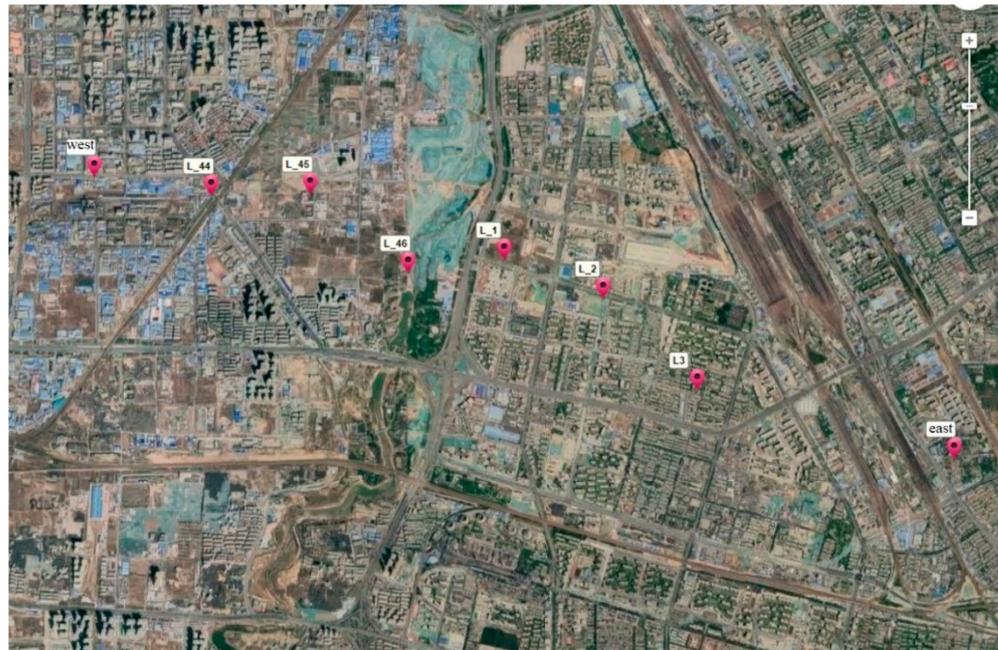


Figure 8. Ground Coverage of Test Field.

Four different calibration approaches were designed to evaluate and compare the proposed quadrifocal tensor SFM photogrammetry method with the original OpenMVG SFM pipeline. The proposed method is tested under different conditions: no Ground Control Point (GCP) support, with GCP support without considering camera intrinsic parameters, and with GCP support considering camera intrinsic parameters.

I. The proposed method with no GCP support

The whole procedure in Figure 5 was implemented using POS. All unknowns, rotation, translation, intrinsic parameters, and 3D ground point coordinates were involved in this process.

II. The proposed method with GCP support, but not considering camera intrinsic parameters

The proposed method with GCP support was evaluated. All the unknowns were employed in the BA procedure except the intrinsic camera parameters. The whole flow was similar to the first test, but the intrinsic parameter matrix remained an identity matrix.

III. The proposed method with GCP support and considering camera intrinsic parameters

The same group of GCPs takes part in the test. All the unknowns, including the intrinsic parameters, are involved in the BA procedure. The intrinsic parameter model adopted Model (4), and the variables α_x and α_y each absorbed two distortion parameters. In the adjustment process, the weight value of the ground control points was 20 times that of the point correspondences.

IV. The original OpenMVG SFM pipeline with GCP support

The same group of GCPs and checkpoints (CHK) was used to evaluate the performance of the original OpenMVG pipeline.

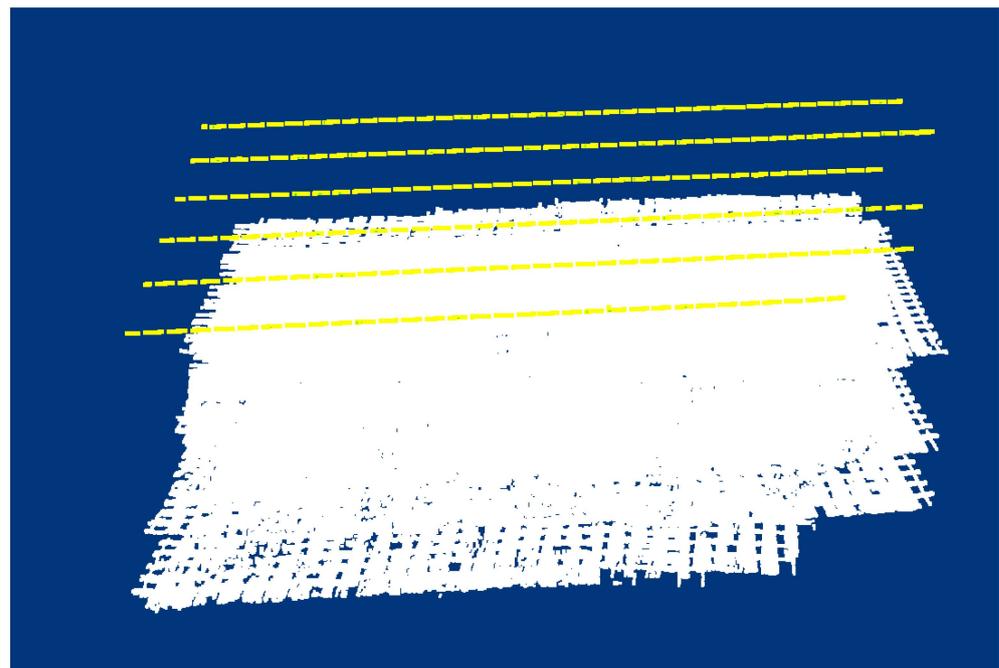
5.2. The Proposed Method with No GCP Support

A total of 7916 images were used in the experiment, and 6,856,633 pairwise point correspondences were extracted from matching. Thirteen ground CHKs were used to measure the discrepancies (i.e., ground point residuals) between the 3D observation values and the calculated estimates. The statistical results of residuals are shown in Table 2, and the root mean squares error (RMSE) is presented in the last row.

Table 2. Statistical Values for CHKs after the First Experiment (Unit: meter).

ID of CHK	X-Axis	Y-Axis	Z-Axis
CHK 1	0.7285	−0.7146	−17.6891
CHK 2	1.3723	−0.5638	−19.4138
CHK 3	0.2392	−0.5449	−16.8526
CHK 4	0.5169	−0.1780	−17.2758
CHK 5	0.2371	−0.5560	−16.6415
CHK 6	−0.0221	−0.3999	−16.5177
CHK 7	1.0317	0.5181	−16.6787
CHK 8	0.6388	0.8845	−16.2159
CHK 9	−1.8655	0.4107	−17.0058
CHK 10	−0.2723	0.1838	−16.7048
CHK 11	0.0234	0.8457	−16.6783
CHK 12	−0.2252	1.3395	−17.2606
CHK 13	−0.6135	0.9441	−17.7322
RMSE	0.7966	0.6950	17.14631

After BA processing, the sparse point cloud of the corresponding area is constructed. The ground point cloud and the corresponding camera projection center are shown in Figure 9. The yellow dashed lines illustrates the position of camera projection centers, and the white form is the ground point cloud.

**Figure 9.** Sparse Point Cloud After the First Experiment.

As presented in Table 1, significant systematic errors were generated when without GCP support, particularly in the Z-axis. From Figure 9, the proposed method can produce a robust sparse, dense cloud.

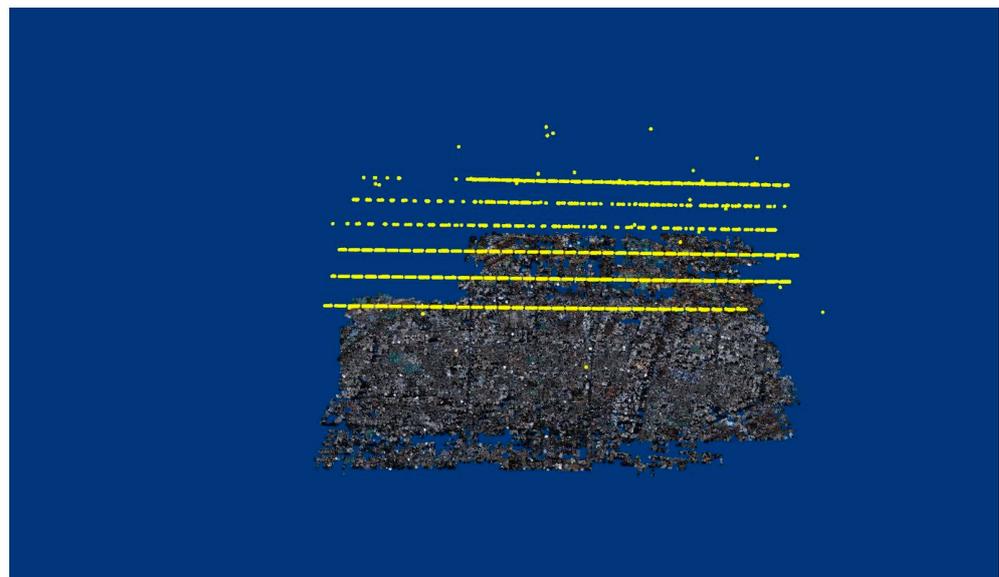
5.3. The Proposed Method with GCP Support, but Not Considering Camera Intrinsic Parameters

The second experiment was carried out with the same set of images. Four ground control points contributed to the BA process, and the same checkpoints were used to evaluate the adjustment accuracy. A total of 3,091,007 pairwise point correspondences were extracted from matching. The statistical differences and RMSE are shown in Table 3.

Table 3. Statistical Values for CHKS after the Second Experiment (Unit: meter).

ID of CHK	X-Axis	Y-Axis	Z-Axis
CHK 1	—	—	—
CHK 2	—	—	—
CHK 3	—	—	—
CHK 4	—	—	—
CHK 5	0.4117	1.0145	−1.6682
CHK 6	0.0062	0.0014	0.0068
CHK 7	—	—	—
CHK 8	—	—	—
CHK 9	−0.0573	0.0133	0.1149
CHK 10	−1.0118	−0.4301	0.0030
CHK 11	−0.0322	−0.0154	−0.0750
CHK 12	—	—	—
CHK 13	−0.0009	0.0054	−0.0014
RMSE	0.3035	0.3057	0.4642

After BA adjustment with GCPs, the pairwise point correspondences were used to construct the sparse ground point cloud, as presented in Figure 10.

**Figure 10.** Sparse Point Cloud After the Second Experiment.

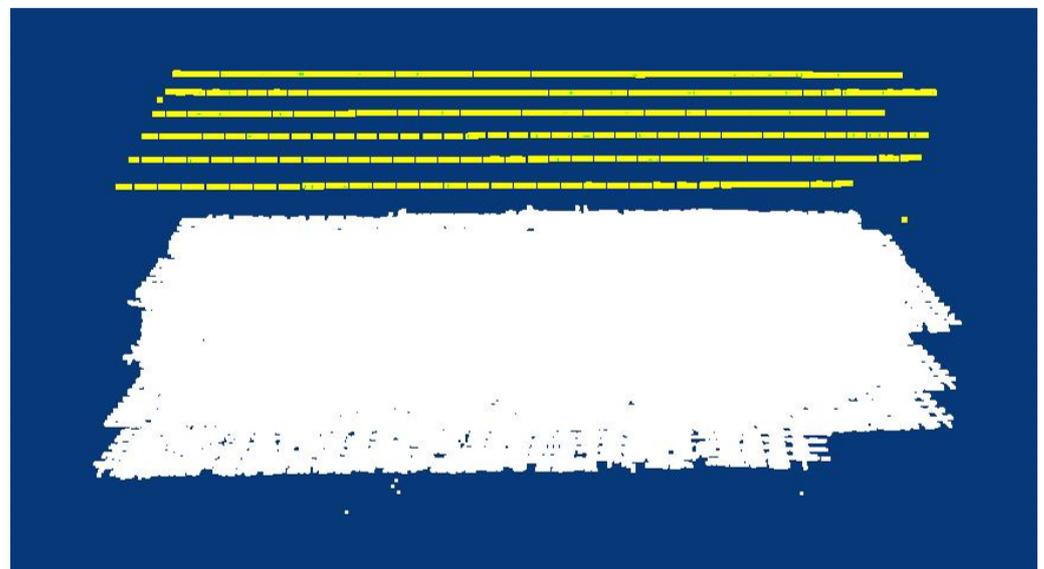
Since the intrinsic camera parameters were not adopted in the BA process, not all distortion sources were fully considered, and some images did not meet the consistency constraint. Therefore, image leaks occurred in the BA processing during the adjustment process; only 4567 images passed through the BA process. Since image leaks transpired in the BA process, some CHKS had no statistical residuals (see Table 3), and the sparse point cloud in Figure 10 lacks completeness and robustness.

5.4. The Proposed Method with GCP Support and Considering Camera Intrinsic Parameters

In the third experiment, a total of 6,884,305 pairwise point correspondences were extracted from matching. The residuals between the 3D observation and calculation values are summarized in Table 4. The sparse point cloud is shown in Figure 11. The intrinsic parameters before and after calibration are presented in Table 5.

Table 4. Statistical Values for CHKS after the Third Experiment.

ID of CHK	X-Axis	Y-Axis	Z-Axis
CHK 1	−0.0081	−0.0001	0.0037
CHK 2	0.0035	0.0034	0.0028
CHK 3	−0.0089	−0.0529	−0.0722
CHK 4	0.0401	−0.0213	−0.0754
CHK 5	0.0571	−0.0411	0.0472
CHK 6	−0.0011	−0.0040	0.0107
CHK 7	−0.0364	−0.0023	−0.0755
CHK 8	−0.0157	−0.0030	−0.0250
CHK 9	−0.1061	0.0232	0.2133
CHK 10	−0.0450	0.0122	0.0258
CHK 11	−0.0034	0.0162	0.0015
CHK 12	0.0076	−0.0027	−0.0047
CHK 13	0.0005	−0.0035	−0.0001
RMSE	0.0392	0.0214	0.0711

**Figure 11.** Sparse Point Cloud After the Third Experiment.**Table 5.** The intrinsic parameters before and after calibration.

Parameter	Before Calibration	After Calibration
Principal Point x_0 (pixel)	2432.0	2423.1399460647876
Principal Point y_0 (pixel)	1616.0	1603.9860816155979
Calibration parameter 1	0	−0.14123671990384698
Calibration parameter 2	0	−27.234841758167391
Calibration parameter 3	0	3807.0519745535798
Calibration parameter 4	0	−0.00018113363239831057
Calibration parameter 5	0	−0.00069940821735464372

Table 3 show that the RMSE values for the 13 CHKS were below one pixel in the X and Y directions and below two pixels in the Z direction, equivalent to the precision obtained by the professional LightSpeed software [12–15]. The sparse point cloud shown in Figure 11 is much more complete and robust.

5.5. The Original OpenMVG SFM Pipeline with GCP Support

The fourth experiment was carried out using the original OpenMVG pipeline. All unknowns were considered, and the same group of GCPs and CHKS were used in the

adjustment. A total of 6,797,129 pairwise point correspondences were extracted from matching. The statistical results are shown in Table 6. The sparse point cloud is displayed in Figure 12.

Table 6. Statistical Values for CHKS after the Fourth Experiment (Unit: meter).

ID of CHK	X-Axis	Y-Axis	Z-Axis
CHK 1	0.2285	−0.2146	0.3560
CHK 2	0.2223	−0.2655	−0.2265
CHK 3	0.2393	−0.2149	−0.2078
CHK 4	0.2362	−0.1163	0.2652
CHK 5	−0.1373	−0.2316	−0.1682
CHK 6	−0.0895	−0.1930	0.3068
CHK 7	0.2187	0.2328	0.1032
CHK 8	0.2167	0.2855	−0.1176
CHK 9	−0.2655	0.2211	0.2149
CHK 10	−0.2353	0.1838	0.2030
CHK 11	0.0215	0.2046	−0.1750
CHK 12	−0.2624	0.2395	0.1582
CHK 13	−0.2297	0.2441	−0.1014
RMSE	0.2998	0.3148	0.3017

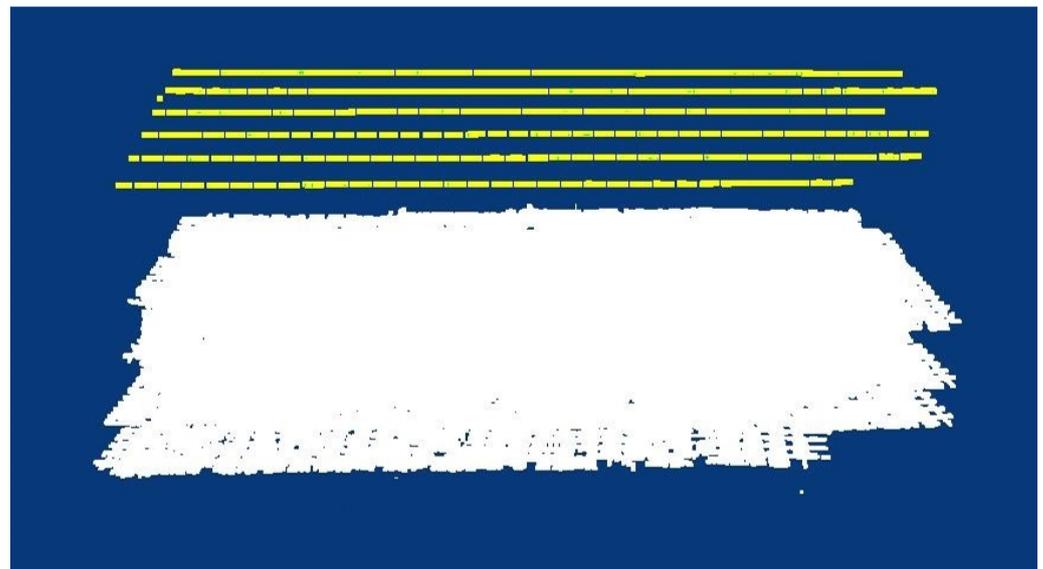


Figure 12. Sparse Point Cloud After the Fourth Experiment.

Table 6 show that the accuracy of the original OpenMVG SFM pipeline is far better than that of the first experiment and that no systematic errors remained in the result. However, the results were still not as good as those of the third experiment.

To better compare and analyze the different test results, the residuals from Tables 2–4 and 6 were plotted. Figure 13 present the residuals at each CHK in the X-direction, Figure 14 show the residuals in the Y-direction, and Figure 15 provide the residuals in the Z-direction. Figure 16 present a more detailed comparison of residuals in the Z-direction.

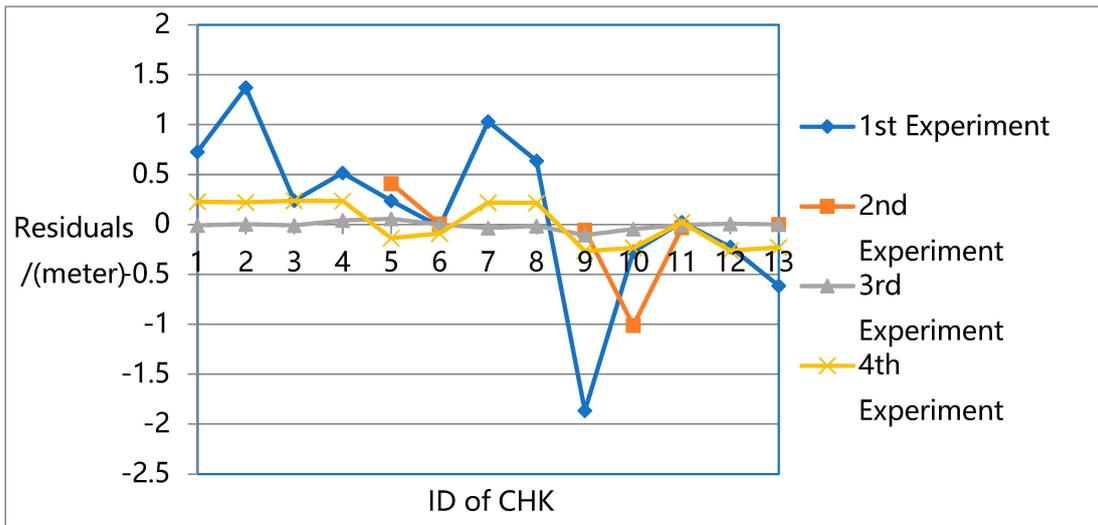


Figure 13. Residuals Plot of CHKs in X-direction.

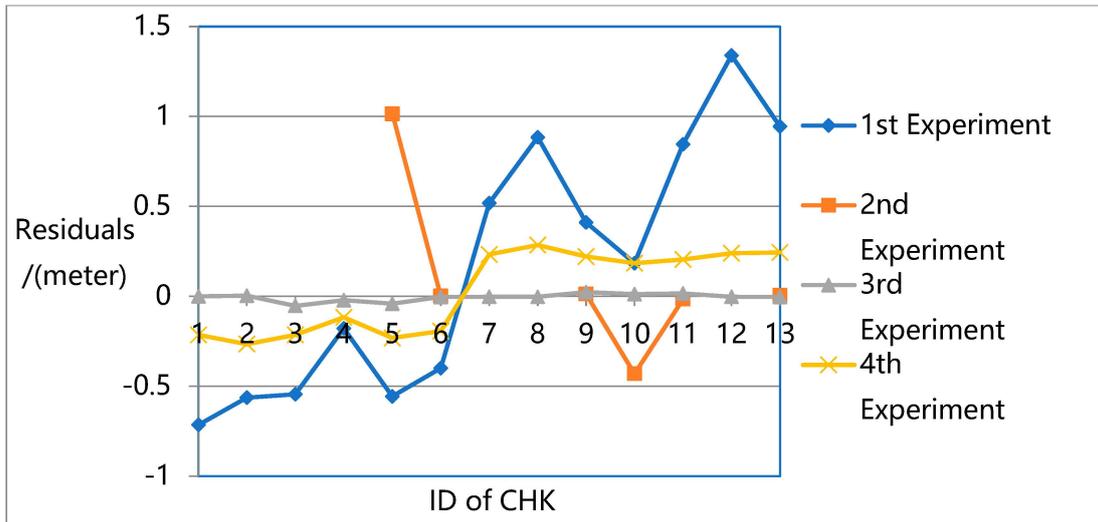


Figure 14. Residuals Plot of CHKs in Y-direction.

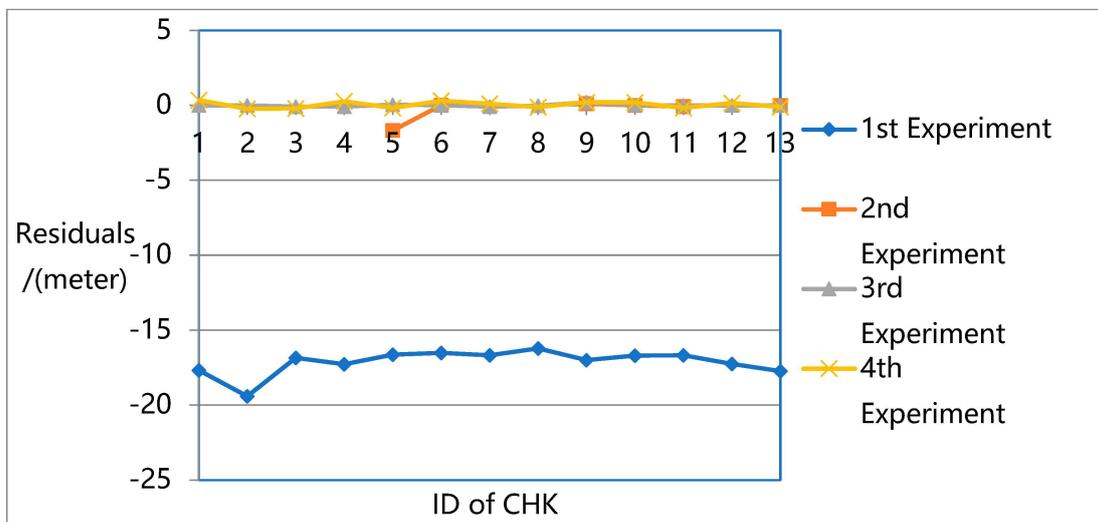


Figure 15. Residuals Plot of CHKs in Z-direction.

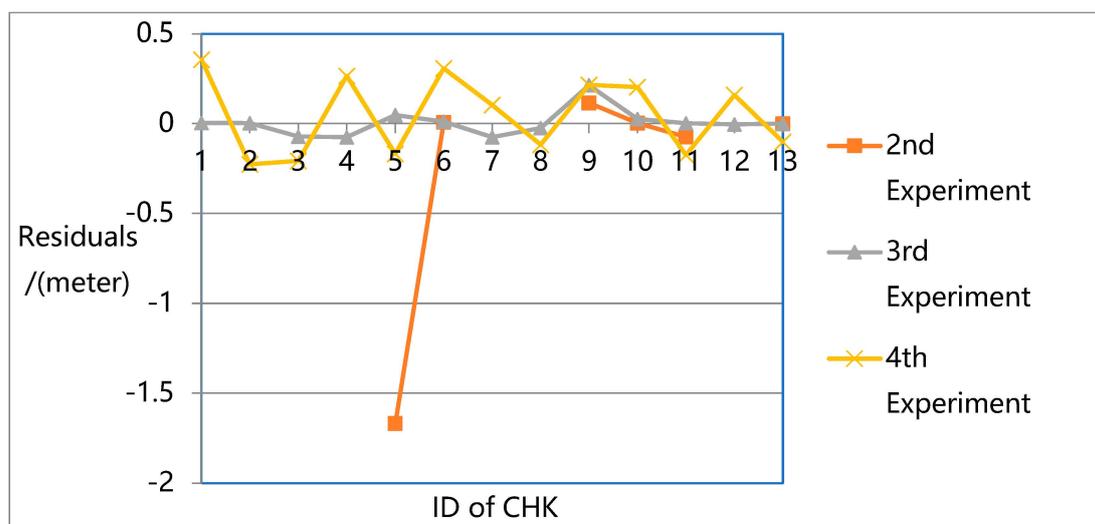


Figure 16. Residuals Plot of CHKs in Z-direction in Detail.

The results show that the largest errors were found in the first experiment. While the first experiment deals with all the unknowns (e.g., rotation, translation, intrinsic parameters, and 3D ground point coordinates) in the overall flow and constructs a robust sparse point cloud, it does not eliminate the systematic errors in the pose data, especially in the Z direction. Due to the lack of ground truth guidance, even with GPS support, significant deviations could be found between the resulting 3D model and the ground truth.

For the second experiment, significant image leaks occurred due to the lack of intrinsic parameters in the adjustment. This means that for high oblique frame sweep images, intrinsic parameters are indispensable in adjustment. For normal vertical photogrammetric images, although the BA process is unable to attain the highest precision when intrinsic parameters are ignored, the robust bundle block can still be obtained.

In the third experiment, the highest precision is reached with a robust sparse point cloud. In the fourth experiment, with GPS and GCP data, the original OpenMVG pipeline can achieve high positioning accuracy in three directions with a robust sparse point cloud. However, its accuracy is not as high as that of the proposed quadrifocal tensor SFM photogrammetry method.

In the third and fourth experiments, while all the unknowns (e.g., rotation, translation, intrinsic parameters, and 3D ground point coordinates) are considered in the BA process, the proposed method achieves much higher accuracy. The proposed approach introduces the iterative photogrammetry idea into the original OpenMVG SFM pipeline following an aerial and satellite photogrammetry processing approach, focusing more on the imaging geometric principle and camera characteristics. Using iterative BA adjustment, the distortion sources are dealt with from a coarse to fine approach, enhancing the BA accuracy step by step.

Figures 17 and 18 show the digital orthophoto map (DOM) and digital elevation model (DEM) products after the BA positioning and calibration process in the third experiment. The ground resolution of the DOM and the DEM was set to 7 cm. Fifteen points were randomly selected from the DOM and DEM products. Measurements were then taken from the two products, and observation values were made in the field test. After comparing the values, the accuracy of the DOM was found to be from 1 to 3 pixels. The largest difference was 2.72 pixels, the smallest difference was 1.01 pixels, and the RMSE was 1.32 pixels. The DEM accuracy ranged from 2 to 5 pixels. The largest difference was 4.23 pixels, the smallest difference was 2.05 pixels, and the RMSE was 2.91 pixels.



Figure 17. DOM Product After Positioning and Calibration.

LightSpeed software was then used for the given dataset, implementing the entire photogrammetry procedure and generating the DOM and DEM products. The photogrammetry results and the DOM and DEM products obtained using the two methods were compared and evaluated. The BA results and the DOM and DEM products obtained by the two approaches were at the same accuracy level, which suggests that our proposed method reaches the precision level of professional photogrammetry software.

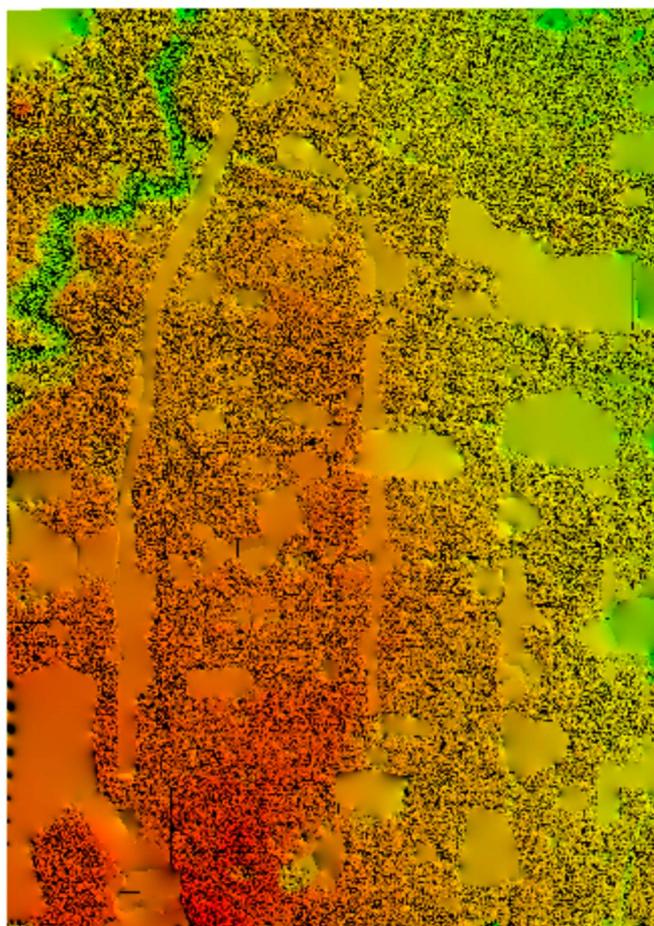


Figure 18. DEM Product After Positioning and Calibration.

6. Conclusions

Sweeping along the flight's vertical direction for imaging, the frame sweep camera is characterized by a large field of view, a wide observation range, and a multi-angle imaging mode. While these characteristics can be useful for 3D reconstruction, they also increase the difficulty of image positioning. For large tilt oblique photogrammetry, the image positioning and calibration process becomes much more complex. Additionally, with the increasing integration between photogrammetry and SFM, more attempts have been made to combine these approaches to realize the positioning, calibration, and 3D reconstruction of large amounts of ordered or disordered images. After an extensive literature review and research, the OpenMVG pipeline was found to be most suitable for VisionMap A3 positioning and calibration.

Using the OpenMVG pipeline, a quadrifocal tensor-based positioning and calibration method was developed for high oblique frame sweep aerial cameras according to the imaging characteristics of VisionMap A3. We comprehensively analyzed the entire processing flow, from feature extraction and image matching, relative rotation and translation estimation, global rotation and translation estimation, and the quadrifocal tensor model construction to the BA process and calibration. Focusing on the imaging character of the VisionMap A3 camera, the quadrifocal tensor was put forward as the basis for BA adjustment. For the BA process and calibration, a coarse to fine three-stage BA processing modification was introduced in the OpenMVG pipeline following photogrammetric processing. Based on the experimental results, our main conclusions are as follows:

First, the SFM photogrammetric approach is suitable for large tilt oblique photogrammetry. When considering the rotation, translation, 3D ground coordinates, and the cam-

era's intrinsic parameters, the SFM pipeline, as OpenMVG, can generate a robust sparse point cloud.

Second, GPS data can only provide an initial value for the BA process. We found significant deviations between the processed bundle block and the ground control points. The results suggest that ground truth is still indispensable in positioning and calibration.

Third, the intrinsic parameters cannot be ignored in the BA process. The lack of intrinsic parameters resulted in image leaks in the bundle block and the sparse point cloud. The results indicate that the intrinsic parameters play an important role in dealing with inner and outer positioning errors.

Fourth, the coarse to fine processing approach in classical photogrammetry is still advisable in SFM photogrammetry. Processing the rotation (R) and translation (T) parameters eliminates most gross positioning errors while processing the intrinsic parameters propels the positioning precision to a higher level.

The experimental results on multiple overlapping routes of VisionMap A3 edge images suggest that our proposed method could achieve a robust bundle block and sparse point cloud and generate accurate DOM and DEM products. However, for the single route, the proposed method generated a bundle block with only contours, and the inner images were mostly considered outliers and excluded from the calculations. Subsequent studies are required to explore how to retain more images and effectively exclude outliers.

Author Contributions: Conceptualization, T.W. and Y.Z. (Yan Zhang); methodology, T.W., Y.Z. (Yan Zhang), and Y.Z. (Yongsheng Zhang); software, T.W. and Y.Z. (Yan Zhang); validation, Y.Y. and L.W.; formal analysis, Z.Z.; writing—original draft preparation, T.W. and Y.Z. (Yan Zhang); writing—review and editing, S.L., L.L. and X.Z.; visualization, Y.Y. and S.L. All authors have read and agreed to the published version of the manuscript.

Funding: This publication has emanated from research conducted with the National Nature Science Foundation of China under grant number 42071340.

Data Availability Statement: The study did not report any data.

Acknowledgments: The authors are grateful to Zhiping Lü and his research team for offering help in processing the GNSS data, which was a preliminary step before we carried out our investigation. Zhiping Lü and his research team have published a free GNSS data process software on www.gnsser.com (accessed on 16 September 2017), and the related paper is <https://link.springer.com/article/10.1007/s10291-020-01069-9> (accessed on 19 January 2021).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript or in the decision to publish the results.

References

1. Konecny, G. *Geoinformation: Remote Sensing, Photogrammetry and Geographic Information Systems*, 2nd ed.; CRC Press: Boca Raton, FL, USA, 2014.
2. Lindner, W. *Digital Photogrammetry: A Practical Course*; Springer: Berlin, Germany, 2016.
3. McGlone, J.C. (Ed.) *Manual of Photogrammetry*, 6th ed.; American Society of Photogrammetry and Remote Sensing: Bethesda, MD, USA, 2013.
4. Wolf, P.R.; Dewitt, B.A. *Elements of Photogrammetry with Applications in GIS*, 4th ed.; McGraw-Hill Education: New York, NY, USA, 2014.
5. Aber, J.S.; Marzolf, I.; Ries, J. *Small-Format Aerial Photography: Principles, Techniques and Geoscience Applications*; Elsevier: Amsterdam, The Netherlands, 2010.
6. Ikkala, L.; Ronkanen, A.K.; Ilmonen, J.; Similä, M.; Rehell, S.; Kumpula, T.; Pääkkilä, L.; Klöve, B.; Marttila, H. Unmanned Aircraft System (UAS) Structure-From-Motion (SfM) for Monitoring the Changed Flow Paths and Wetness in Minerotrophic Peatland Restoration. *Remote Sens.* **2022**, *14*, 3169. [[CrossRef](#)]
7. Hinz, A.; Dörstel, C.; Heier, H. *DMC-The digital sensor technology of Z/I-Imaging. Photogrammetric Week'01*; Fritsch, D., Spiller, R., Eds.; Wichmann Verlag: Heidelberg, Germany, 2001.
8. Dhillon, V.S.; Marsh, T.R.; Stevenson, M.J.; Atkinson, D.C.; Kerry, P.; Peacocke, P.T.; Vick, A.J.A.; Beard, S.M.; Ives, D.J.; Lunney, D.W.; et al. ULTRACAM: An ultrafast, triple-beam CCD camera for high-speed astrophysics. *Mon. Not. R. Astron. Soc.* **2007**, *378*, 825–840. [[CrossRef](#)]

9. Toschi, I.; Remondino, F.; Rothe, R.; Klimek, K. Combining Airborne Oblique Camera and Lidar Sensors: Investigation and new Perspectives. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, *42*, 437–444. [[CrossRef](#)]
10. Li, J.; Liu, X.; Wan, Y.; Liu, F. Generation of SWDC-4 aerial digital camera virtual images. *Geomat. Inf. Sci. Wuhan Univ.* **2008**, *33*, 450–453.
11. Hoeffft, J.R.; Tietz, T. Multispectral EO LOROP camera. In *Infrared Technology and Applications XXIV*; SPIE: Bellingham, WA, USA, 1998; Volume 3436, pp. 402–412.
12. Iyengar, M.; Lange, D. The Goodrich 3rd generation DB-110 system: Operational on tactical and unmanned aircraft. In *Airborne Intelligence, Surveillance, Reconnaissance (ISR) Systems and Applications III*; SPIE: Bellingham, WA, USA, 2006; Volume 6209, pp. 75–88.
13. Pechatnikov, M.; Shor, E.; Raizman, Y. VisionMap A3-super wide angle mapping system basic principles and workflow. In Proceedings of the 21th ISPRS Congress, Beijing, Oral Technical Session SS-8 (2), Beijing, China, 3–11 July 2008.
14. Raizman, Y.; Gozes, A. *High Throughput Aerial Photography, Ortho & 3D Processing*. *Photogrammetric Week*; Fritsch, D., Spiller, R., Eds.; Wichmann Verlag: Heidelberg, Germany, 2013.
15. Raizman, Y.; Gozes, A. Tel-Aviv. VisionMap A3 Edge—A Single Camera for Multiple Solutions. In Proceedings of the Conference 55th Photogrammetric Week, Stuttgart, Germany, 7–11 September 2015.
16. Pechatnikov, M.; Raizman, Y.; Tel-Aviv, V. Visionmap A3 LightSpeed—A new generation of the ground processing system. In Proceedings of the International Scientific Racurs Conference, Tossa der Mar, Spain, 19–22 September 2011.
17. Triggs, B.; McLauchlan, P.F.; Hartley, R.I.; Fitzgibbon, A.W. Bundle adjustment—A modern synthesis. In *International Workshop on Vision Algorithms*; Springer: Berlin/Heidelberg, Germany, 1999; pp. 298–372.
18. Richard, H.; Andrew, Z. *Multiple View Geometry in Computer Vision*; Cambridge University Press: Cambridge, UK, 2000.
19. Khan, P.W.; Byun, Y.C.; Latif, M.A. Clifford Geometric Algebra-Based Approach for 3D Modeling of Agricultural Images Acquired by UAVs. *IEEE Access* **2020**, *8*, 226297–226308. [[CrossRef](#)]
20. Zhang, D.; Jackson, W.; Dobie, G.; West, G.; MacLeod, C. Structure-from-motion based image unwrapping and stitching for small bore pipe inspections. *Comput. Ind.* **2022**, *139*, 103664. [[CrossRef](#)]
21. Jiang, N.; Cui, Z.; Tan, P. A global linear method for camera pose registration. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, NSW, Australia, 1–8 December 2013; pp. 481–488.
22. Cui, Z.; Tan, P. Global structure-from-motion by similarity averaging. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 864–872.
23. Cui, Z.; Jiang, N.; Tang, C.; Tan, P. Linear global translation estimation with feature tracks. *arXiv* **2015**, arXiv:1503.01832.
24. Goldstein, T.; Hand, P.; Lee, C.; Voroninski, V.; Soatto, S. Shapefit and shapekick for robust, scalable structure from motion. In Proceedings of the European Conference on Computer Vision; Springer: Cham, Switzerland, 2016; pp. 289–304.
25. Sweeney, C.; Sattler, T.; Hollerer, T.; Turk, M.; Pollefeys, M. Optimizing the viewing graph for structure-from-motion. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 801–809.
26. Ozyesil, O.; Singer, A. Robust camera location estimation by convex programming. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 2674–2683.
27. Wilson, K.; Snavely, N. Robust global translations with 1dsfm. In Proceedings of the European Conference on Computer Vision; Springer: Cham, Switzerland, 2014; pp. 61–75.
28. Zhu, S.; Zhang, R.; Zhou, L.; Shen, T.; Fang, T.; Tan, P.; Quan, L. Very large-scale global sfm by distributed motion averaging. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4568–4577.
29. Moulon, P.; Monasse, P.; Marlet, R. Adaptive structure from motion with a contrario model estimation. In Proceedings of the Asian Conference on Computer Vision; Springer: Berlin/Heidelberg, Germany, 2012; pp. 257–270.
30. Cui, H.; Shen, S.; Gao, X.; Hu, Z. Batched incremental structure-from-motion. In Proceedings of the International Conference on 3D Vision (3DV), Qingdao, China, 10–12 October 2017; pp. 205–214.
31. Snavely, N.; Seitz, S.M.; Szeliski, R. Modeling the world from internet photo collections. *Int. J. Comput. Vis.* **2008**, *80*, 189–210. [[CrossRef](#)]
32. Snavely, N.; Seitz, S.M.; Szeliski, R. Photo tourism: Exploring photo collections in 3D. In *ACM Siggraph 2006 Papers*; ACM: Boston, MA, USA, 2006; pp. 835–846.
33. Cui, H.; Shen, S.; Gao, W.; Wang, Z. Progressive large-scale structure-from-motion with orthogonal MSTs. In Proceedings of the 2018 International Conference on 3D Vision (3DV), Verona, Italy, 5–8 September 2018; pp. 79–88.
34. Schonberger, J.L.; Frahm, J.M. Structure-from-motion revisited. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 4104–4113.
35. Wu, C. Towards linear-time incremental structure from motion. In Proceedings of the 2013 International Conference on 3D Vision-3DV 2013, Seattle, WA, USA, 29 June–1 July 2013; pp. 127–134.
36. Moulon, P.; Monasse, P.; Marlet, R. Global fusion of relative motions for robust, accurate and scalable structure from motion. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, NSW, Australia, 1–8 December 2013; pp. 3248–3255.

37. Lindenberger, P.; Sarlin, P.E.; Larsson, V.; Pollefeys, M. Pixel-Perfect Structure-from-Motion with Feature metric Refinement. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 5987–5997.
38. Bhowmick, B.; Patra, S.; Chatterjee, A.; Govindu, V.M.; Banerjee, S. Divide and conquer: Efficient large-scale structure from motion using graph partitioning. In *Asian Conference on Computer Vision*; Springer: Cham, Switzerland, 2014; pp. 273–287.
39. Gherardi, R.; Farenzena, M.; Fusiello, A. Improving the efficiency of hierarchical structure-and-motion. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 1594–1600.
40. Farenzena, M.; Fusiello, A.; Gherardi, R. Structure-and-motion pipeline on a hierarchical cluster tree. In Proceedings of the 2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops, Kyoto, Japan, 27 September–4 October 2009; pp. 1489–1496.
41. Toldo, R.; Gherardi, R.; Farenzena, M.; Fusiello, A. Hierarchical structure-and-motion recovery from uncalibrated images. *Comput. Vis. Image Underst.* **2015**, *140*, 127–143. [[CrossRef](#)]
42. Cui, H.; Gao, X.; Shen, S.; Hu, Z. HSfM: Hybrid structure-from-motion. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1212–1221.
43. Zhu, S.; Shen, T.; Zhou, L.; Zhang, R.; Wang, J.; Fang, T.; Quan, L. Parallel structure from motion from local increment to global averaging. *arXiv* **2017**, arXiv:1702.08601.
44. Shah, R.; Deshpande, A.; Narayanan, P.J. Multistage SfM: A coarse-to-fine approach for 3d reconstruction. *arXiv* **2015**, arXiv:1512.06235.
45. Bao, S.Y.; Savarese, S. Semantic structure from motion. In *CVPR 2011*; IEEE: Colorado Springs, CO, USA, 2011; pp. 2025–2032.
46. Bao, S.Y.; Bagra, M.; Chao, Y.W.; Savarese, S. Semantic structure from motion with points, regions, and objects. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 2703–2710.
47. Chen, Y.; Wang, Y.; Lu, P.; Chen, Y.; Wang, G. Large-scale structure from motion with semantic constraints of aerial images. In *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*; Springer: Cham, Switzerland, 2018; pp. 347–359.
48. Westoby, M.J.; Brasington, J.; Glasser, N.F.; Hambrey, M.J.; Reynolds, J.M. ‘Structure-from-Motion’ photogrammetry: A low-cost, effective tool for geoscience applications. *Geomorphology* **2012**, *179*, 300–314. [[CrossRef](#)]
49. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *IJCV* **2004**, *60*, 91–110. [[CrossRef](#)]
50. Martinec, D.; Pajdla, T. Robust rotation and translation estimation in multi-view reconstruction. In Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17–22 June 2007; pp. 1–8.
51. Wang, T.; Zhang, Y.; Zhang, Y.; Zhang, Z.; Xiao, X.; Yu, Y.; Wang, L. A Spliced Satellite Optical Camera Geometric Calibration Method Based on Inter-Chip Geometry Constraints. *Remote Sens.* **2021**, *13*, 2832. [[CrossRef](#)]
52. Wang, T.; Zhang, Y.; Zhang, Y.; Jiang, G.; Zhang, Z.; Yu, Y.; Dou, L. Geometric Calibration for the Aerial Line Scanning Camera GFXJ. *Photogramm. Eng. Remote Sens.* **2019**, *85*, 643–658. [[CrossRef](#)]