

Automatic Sensor Orientation using Horizontal and Vertical Line Feature Constraints

Yanbiao Sun^{a,b,*}, Stuart Robson^b, Daniel Scott^b, Jan Boehm^b, Qiang Wang^c

^a*State Key Laboratory of Precision Measurement Technology and Instruments, Tianjin University, Tianjin, 300072, China*

^b*Department of Civil, Environmental & Geomatic Engineering, University College London, London, W1W6BT, UK*

^c*College of Geoscience and Surveying Engineering, China University of Mining and Technology, Beijing, China*

Abstract

To improve the accuracy of sensor orientation using calibrated aerial images, this paper proposes an automatic sensor orientation method utilizing horizontal and vertical constraints on human-engineered structures, addressing the limitations faced with sub-optimal number of Ground Control Points (GCPs) within a scene. Related state-of-the-art methods rely on structured building edges, and necessitate manual identification of end points. Our method makes use of line-segments but eliminates the need for these matched end points, thus eliminating the need for inefficient manual intervention.

To achieve this, a 3D line in object space is represented by the intersection of two planes going through two camera centers. The normal vector of each plane can be written as a function of a pair of azimuth and elevation angles. The normal vector of the 3D line can be expressed by the cross product of these two plane's normal vectors. Then, we create observation functions of horizontal and vertical line constraints based on the zero-vector cross-product and the dot-product of the normal vector of the 3D lines. The observation functions of the horizontal and vertical lines are then introduced into a hybrid Bundle Adjustment (BA) method as constraints, including observed image points as well as observed line segment projections. Finally,

*Corresponding author

Email addresses: Yanbiao.Sun@ucl.ac.uk (Yanbiao Sun), s.robson@ucl.ac.uk (Stuart Robson), daniel.scott.12@ucl.ac.uk (Daniel Scott), j.boehm@ucl.ac.uk (Jan Boehm), wangqiangstudy@163.com (Qiang Wang)

to assess the feasibility and effectiveness of the proposed method, simulated and real data are tested. The results demonstrate that, in cases with only 3 GCPs, the accuracy of the proposed method utilizing line features extracted automatically, is increased by 50%, compared to a BA using only point constraints.

Keywords: sensor orientation, line features, bundle adjustment, horizontal lines, vertical lines.

1. Introduction

One of the central tasks within aerial photogrammetry is simultaneous determination of a sensor's Exterior Orientation Parameters (EOPs) and the 3D position of observed points in object space. Knowledge of a sensor's orientation can then be used as the basis for dense matching, ortho-photo generation and so on (Gerke, 2009; James and Robson, 2012; Zhang et al., 2012; Gülch, 2012), and thus the accuracy of sensor orientation strongly impacts on that of latter photogrammetric procedures.

To determine the EOPs of only one image, a resection is an effective method when more than three Ground Control Points (GCPs) and corresponding image points are available. Similarly, a relative orientation method can be utilized to estimate the EOPs of stereo images (Zhang et al., 2011). However, in the majority of cases it is necessary to compute the EOPs of many aerial images distributed on multiple tracks. Bundle Adjustment (BA) is regarded as a most efficient and economic method to achieve accurate sensor orientation. For the traditional BA method, many GCPs distributed evenly across the surveyed area are required to reduce any propagation of errors across regions. For example, in (Alamus et al., 2006) 42 GCPs were established in order to provide EOPs for 390 DMC images taken at an average altitude of 4500 m above the ground level. Establishing such GCPs can be a labor-intensive and costly process.

To reduce the number of GCPs required for a survey, common practice is to integrate high-precision and expensive GNSS/IMU sensors into the photogrammetric camera system. For example, in (Yuan et al., 2009) high-accuracy precise point positioning (PPP) GNSS data served as an additional observation in their BA in order to reduce the number of the required GCPs. With four GCPs situated at the extents of the surveyed region, together with data from two additional cross flights, the accuracy of the method was still

found to be slightly inferior to that of a traditional BA with complete GCP coverage (the accuracy was, however, sufficient for the topographic mapping requirements for which the approach was intended). In addition, when both high-precision Inertial Measurement System (IMU) and GPS are present, a direct sensor orientation can be used to extrapolate from the projections centers to the ground coordinate system (Yastikli and Jacobsen, 2005).

However, in many cases such as many low-cost or Unmanned Aerial Vehicle (UAV) applications, GPS/IMU instruments are unavailable, or their precision is poor, rendering a GPS/IMU augmented BA approach infeasible. Thus, during the past two decades BA methods with additional sensor data or constraints have been proposed. For example, when multi-temporal aerial imageries covering the same area were available, previous world point solutions were input into a combined BA method to achieve more robust orientation of new aerial images (Yuan et al., 2004). To address the limitation of the traditional BA method applied in weak convergence geometry, DEM-aided BA method was proposed by (Teo et al., 2010), in which Digital Elevation Model (DEM) was used as the elevation control. To overcome the need for GCPs required by the traditional BA method, a dense network of reference points were extracted by a least squares surface matching and then used for GCPs (Gneeniss et al., 2013).

In close-range photogrammetric applications, researchers in computer/robotic communities have proposed many optimization models based on both point and line features so as to improve the robustness in both textured and structured environments. (Jeong et al., 2006) used both line and corner features as landmarks in a new vision-based SLAM, and an extended Kalman filter-based framework was utilized for localization and reconstruction. (Ramalingam et al., 2011) developed a general technique that can solve minimal pose estimation configurations using points and lines by making full use of collinearity and coplanarity constraints. In order to overcome failure of Perspective-n-Point (PnP) problem in low texture environment, (Vakhitov et al., 2016) introduced an algebraic line error formulated as linear constraints on the line endpoints so as to handle points, lines, or a combination of them. (Zuo et al., 2017) employed orthonormal representation as the minimal parameterization of modelling line features and then developed robust efficient visual SLAM system based heterogeneous point and line features, significantly improving the SLAM solution. (Hofer et al., 2017) utilized both straight line and point feature to achieve the task of 3D scene abstraction so as to overcome the high computational complexity associated with subsequent multi-view stere-

o approaches. (Gomez-Ojeda et al., 2017) developed a stereo visual SLAM system (PL-SLAM), which combined both points and line segments to work robustly in a wider variety of scenarios, particularly in low-textured environments. (Zheng et al., 2018) presented a tightly-coupled filtering-based stereo VIO system using both points and lines, in which line features were also used to help improve system robustness when point features cannot be reliably detected. From the experimental results shown in the above published work, we note that line features used in each optimization model were introduced mainly for the sake of improving the robustness, especially in low-texture environments, although modest improvements in accuracy were a welcomed side effect.

In aerial photogrammetric application, (Gerke, 2011) presented a state-of-the-art method to integrate linear horizontal, vertical and right-angled scene structures into the BA method, predicated on the assumption that within a given scene most “straight lines” are going to be of engineered human origin, especially within industrial or urban environments. Where objects are of human-origin, incentives for features to be straight, plumb or level, are myriad: be it due to the nature of contemporary engineering design and construction practices, sociological issues around land ownership and city planning, gravitational loading on structures, or the fundamental geometric properties of the line in 3D space. His experimental results demonstrated the number of GCPs necessary can be successfully reduced when including these assumptions. However, although the scene structures used are ostensibly line features, the two end-points of each feature need to be identified by hand to define the constraint. For horizontal lines, the z-components of two end-points were identical, while the xy-components were the same for vertical lines. Meanwhile, right-angled constraints were defined by a 90^0 triangle consisting of three corners.

We find that the horizontal and vertical scene structures used in (Gerke, 2011) can be expressed by line features without end-points. But, the parameterization of line features is more complicated than that of point features, and the BA method created on basis of line parameterization is also complicated. Fortunately, geometric computation based on line features is well established within the photogrammetric community. A simple representation of a 3D line segment is two end-points along the object-space straight line (Habib et al., 2002). The standard representation used in photogrammetry is point-orientation definition (Schenk, 2004), in which a fixed point and a direction vector jointly expressed a line. As a line in 3D has only 4 Degrees

of Freedom (DoF), the method required two necessary constraints; that the fixed point was the closest to the Origin, and that sign ambiguity was imposed onto the lines direction. Another well-known representation is Plücker matrices, in the form of a 4×4 skew-symmetric homogenous matrix having 5 dependent ratios (Hartley and Zisserman, 2003). Because the representation was over-parameterized, one constraint of zero determinate of the matrix was imposed. (Roberts, 1988) proposed a four-parameter line parameterization with two orientation parameters and two positional parameters, in which the positional parameters were defined by the intersection between the line and XY plane. Recently, (Zhao et al., 2015) had proposed a new monocular SLAM algorithm using straight lines, in which a 3D line was defined by two pairs of azimuth and elevation angles. Here, the normal vector of the plane passing through a camera center was expressed by an azimuth and elevation angles. Thus, two planes passing through two corresponding line projections can define a 3D line by the plane intersection. The results of the method exploited in SLAM have demonstrated to achieve accurate and robust estimation, thus in this study, the parameterization of (Zhao et al., 2015) is used in the proposed method.

Thus, to address the limitation of manual work required by (Gerke, 2011) this study eliminates the need for manual measurements of end-points of line structures by extrapolating line segments into lines (of infinite length). Structures in image space are represented using 2D infinite line features, while corresponding 3D line in object space are represented using two plane normal vectors. Thus, the normal vector of the structure lines can be calculated by the cross product of these two planes. When the lines are horizontal, the cross product between their normal vectors and the Z-axis on a geodetic coordinate system should be zero. Similarly, for vertical lines, the dot product between their normal vectors and the Z-axis is also zero. The horizontal and vertical lines are used for additional observations and input to a hybrid BA model, which consists of two observed features, namely, image points and 2D line features.

The rest of this article is organized as follows. In Section 2, we briefly introduce the representation of a 3D line and the observation function with which the line in object space can project into line projection in image space. After that, the horizontal and vertical line feature constraints are defined by the cross product and dot product between Z-axis and line normal vectors in the geodetic coordinate system. Section 3 describes a Hybrid BA (HBA) method, integrating image point features and 2D line features. Section 4

gives a suggested workflow for the method, including feature extraction and matching, a free-net BA model as well as the HBA model. Section 5 uses synthetic and real datasets to evaluate the accuracy of the proposed method. Section 6 discusses the behavior of the approach with respect to changes in constraints and observations. Finally, we draw our conclusions in Section 7.

2. Line Features Constraints

2.1. 3D line feature representation

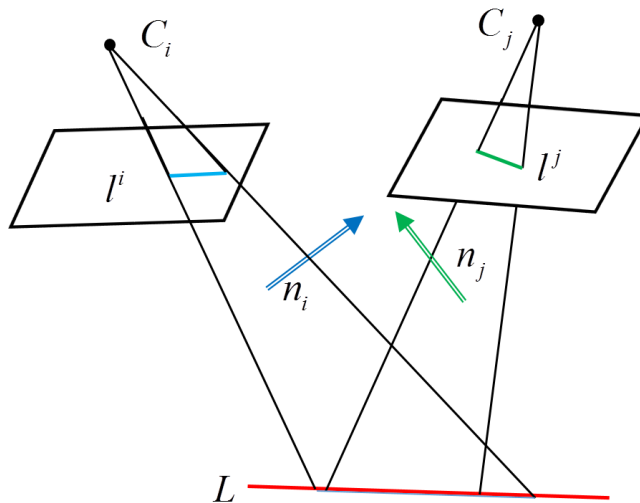


Figure 1: A 3D line feature representation using 4 directional angles, defining two plane normal vectors.

In Figure 1, a red 3D line, L , projects into two cameras, C_i and C_j , forming two line projections, namely, l^i and l^j , respectively. n_i denotes the normal vector of the plane, which passes through center of the C_i camera and the 3D line. Similarly, the normal vector of the plane passing through the center of the C_j camera is n_j . The 3×1 plane normal vector can be computed via

$$n = R^T K^T l \quad (1)$$

where K and R represent a camera calibration matrix and a rotation matrix, respectively. In addition, a direction vector (normal vector) illustrated in

Figure 2 can be jointly represented using an azimuth angle (φ) and elevation (θ) in

$$n = N(\varphi, \theta) = \begin{bmatrix} \sin\theta\cos\varphi \\ \sin\theta\sin\varphi \\ \cos\theta \end{bmatrix} \quad (2)$$

A 3D line can be defined by the intersection of two planes. Thus, in this paper, two normal vectors, including two pairs of azimuth and elevation angles, are used to represent a 3D line, namely

$$L = [\varphi_i, \theta_i, \varphi_j, \theta_j] \quad (3)$$

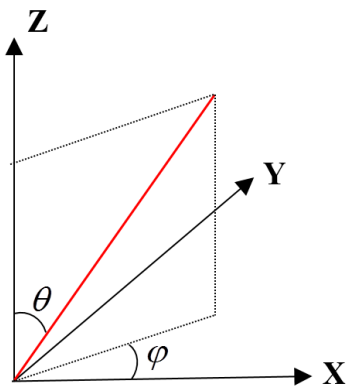


Figure 2: A direction vector represented by an azimuth angle (φ) and elevation (θ).

2.2. Observation function of 3D lines

In Figure 3, the red 3D line L also projects onto a third image, C_k , forming a third line projection, l^k . Since a 3D line can be just defined by two plane normal vectors, no redundant constraints are provided if only two line projections can be observed. 3D lines with more than two line projections provide line projection constraints based on the trifocal tensor model. As stated in Section 2.1, a 3D line is defined by two plane normal vectors with two pairs of azimuth and elevation angles. The third normal vector of the plane passing through can be calculated based on the trifocal tensor model (Hartley and Zisserman, 2003) via

$$n_k = (t_j - t_k)^T n_j n_i - (t_i - t_k)^T n_i n_j \quad (4)$$

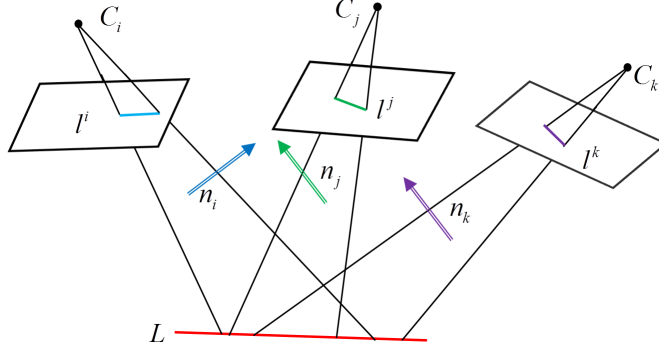


Figure 3: The line feature constraints from more two images can be established based on trifocal tensor model.

where t_i , t_j and t_k represent the locations of the i^{th} , j^{th} , and k^{th} cameras, respectively. Then, the projections of lines can be obtained by an observation function of

$$l^\Psi = K^{-T} R n_\Psi, \Psi = \{i, j, k\} \quad (5)$$

building a relationship between a 3D line and a 2D line segment.

The cost function of the line projections can be defined in Equation (6) to minimize the Euclidean distances between the predicted 2D line projections and edge points of the observed 2D line segments.

$$\arg \min (x_i^T l^\Psi)^T (x_i^T l^\Psi) \quad (6)$$

where a 2D line segment is composed with lots of edge points, namely, $l = \{x_1, x_2, \dots, x_i, \dots, x_M\}$. Thus, a 2D line segment can provide 3 observation equations, namely

$$g(X) : \arg \min l^{\Psi T} \Pi l^\Psi \quad (7)$$

where $\Pi = \sum x_i x_i^T$ is a 3×3 matrix.

2.3. Horizontal and vertical lines constraints

As explained in Section 2.1, a 3D line, L , can be represented using two plane normal vectors (n_M and n_A), namely

$$n_M = \begin{bmatrix} \sin\theta_M \cos\varphi_M \\ \sin\theta_M \sin\psi_M \\ \cos\theta_M \end{bmatrix}, n_A = \begin{bmatrix} \sin\theta_A \cos\varphi_A \\ \sin\theta_A \sin\psi_A \\ \cos\theta_A \end{bmatrix} \quad (8)$$

Here (φ_M, θ_M) and (φ_A, θ_A) are two pairs of azimuth and elevation angles on the M^{th} and A^{th} cameras, respectively (M and A denote the main and associate camera, respectively. Two cameras with biggest parallax angle are selected as the main and associate cameras.). Therefore, the directional vector of the 3D line can be calculated by the cross product of these two plane normal vectors, namely,

$$n_L = n_M \times n_A \quad (9)$$

Meanwhile, the Z-axis vector in a geodetic coordinate system can be written by

$$n_V = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad (10)$$

Therefore, if a 3D line is horizontal, the dot product between the L and the Z-axis vector should be zero, namely,

$$h(X) = (n_M \times n_A)^T \cdot n_V, h = 0 \quad (11)$$

For a vertical 3D line, the cross product between L and the Z-axis is a 3×1 zero vector and computed by

$$v(X) = (n_M \times n_A)^T \times n_V, v = 0 \quad (12)$$

For above two equations, X represent the unknown parameters, including EOPs and 3D line parameters. In this paper, the zero-vector dot product and zero-vector cross product are used as the constraint on the horizontal and vertical lines, and then imposed into the cost function explained below.

3. Hybrid Bundle Adjustment using Point and Line Observations

In this study, both image points and line segments are treated as measured observations. Our approach, using these two types of features, is thus termed HBA.

A 3D point in object space, M , projects into a camera forming an image point where

$$f = \begin{bmatrix} u \\ w \end{bmatrix} = \begin{bmatrix} x/z \\ y/z \end{bmatrix} \quad (13)$$

where

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = KR(M - t) \quad (14)$$

Here, u and w denote x- and y-direction locations of an image point respectively. An image point is also represented in the form of a homogenous vector $[x, y, z]^T$.

Thus, the HBA's cost function can be expressed by

$$\arg \min_X \rho_f \|f(X) - Z\|_{\Sigma_f^{-1}}^2 + \rho_g \|g(X) - \mathbf{0}\|_{\Sigma_L^{-1}}^2 + \rho_h \|h(X) - \mathbf{0}\|_{\Sigma_h^{-1}}^2 + \rho_v \|v(X) - \mathbf{0}\|_{\Sigma_v^{-1}}^2 \quad (15)$$

where Z denotes measured image point projections. ρ_f, ρ_g, ρ_h and ρ_v are robust Huber cost functions (Zuo et al., 2017). $\Sigma_f^{-1}, \Sigma_L^{-1}, \Sigma_h^{-1}$ and Σ_v^{-1} denote the inverse matrices of the covariance matrices of four observations, namely, image points, image line segments, horizontal constraint and vertical constraints, respectively. The four weighting matrices are manually set, namely, $\Sigma_h^{-1} = \Sigma_v^{-1} = \lambda \Sigma_f^{-1} = \lambda \Sigma_L^{-1}$. λ is called as weighting ratio parameters and will be discussed in Section 6. For the cost function, the unknown parameters include camera pose (C), 3D points (M) and 3D line (L), namely,

$$X = [C \quad M \quad L]^T \quad (16)$$

The structure of the HBA's Jacobian matrix is shown in Figure 4, including the four kinds of observation, namely; Image Points, Line Segments, Horizontal Constraints and Vertical Constraints. Assume that N_P image points (\aleph_P

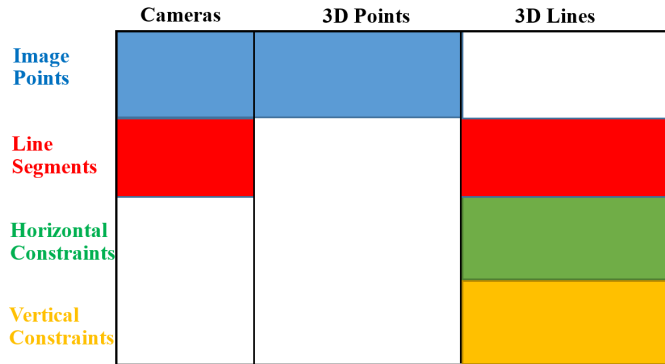


Figure 4: Jacobian structure of HBA.

3D points) and N_L line segments (\aleph_L 3D lines) are extracted from \aleph_C images,

among which N_H and N_V 3D lines are horizontal and vertical, respectively. Although we add horizontal and vertical lines constraints, the unknown parameters is not changed, and the number of those is $6N_C + 3N_P + 4N_L$. The number of the observations is $2N_P + 3N_L + N_H + 2N_V$.

When all of the unknown parameters are initialized, we can iteratively search the descent path to a minimum based on Gauss-Newton or Levenberg-Marquardt optimization solutions (Sun et al., 2016).

4. Workflow

To implement this approach, we sequentially conduct three main procedures, shown in Figure 5, namely; Feature Extraction & Matching (1), Free-net BA Optimization (2), and the HBA method with line constraints (3).

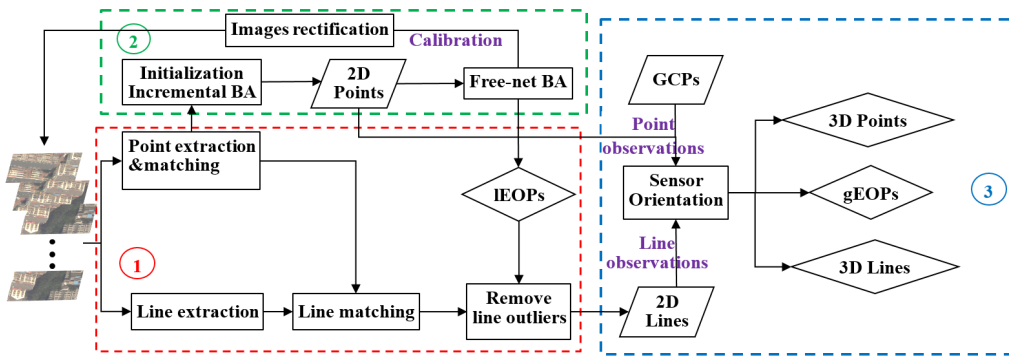


Figure 5: Workflow of the proposed method, including three main procedures.

4.1. Feature Extraction & Matching

The proposed method includes two types of features; image interest points and line segments. Thus, the first step of the proposed method must be the extraction and matching of image points and line segments. In this study, we extract high-quality image point correspondences using a publicly software package (L2-SIFT). The package uses a block-SIFT algorithm to extract points from large-size aerial images. A red-black data structure is then used to efficiently achieve matching image points from multiple images (Sun et al., 2014). Here, a two-level RANSAC algorithm is first used to obtain as many point matches from two images as possible; an incremental BA algorithm

is then used to further remove incorrect point matches for additional images. “CannyLines”, a parameter-free Canny operator to robustly extract edge map, is used to extract line segments from an input image (Lu et al., 2015). Where lines are of less than 100 pixels in length, they are excluded due to unreliability of short line segments. Once image points are matched, we can accomplish stereo matching of line segments based on point-line invariance (Fan et al., 2012). Finally, we remove line and point outliers based on their projective residuals, where these are indicative of incorrect line segment matches.

As a line in 3D has 4 degrees of freedom, (instead of 3 for a point), more than 4 observations are required to distinguish line outliers. Thus, lines appearing on less than 3 images are rejected. Remaining line matches are checked based on the calculated line projection residual error.

Given the i^{th} set of line segment matches $L_i = \{l_i^a, l_i^b, l_i^c\}$ (where a, b and c denote the index of cameras), we can take any two line segments (say, l_i^a and l_i^b) to determine an unbounded 3D line say \hat{l}_i . We can then project \hat{l}_i onto the rest of the images, obtaining a predicted infinite 2D line, say \hat{l}_i^c (for camera c). The observed line segment with the length of n pixels on c^{th} image is firstly divided into n sampled edge points, namely, $l_i^c = \{p_i^1, p_i^2, \dots, p_i^n\}$. Then, the average normed distance between these edge points of l_i^c and \hat{l}_i is

$$d_L = \frac{1}{n} \sum \|p_i^j{}^T \hat{l}_i\| \quad (17)$$

When d_L is more than a given threshold (1 pixel is set in this paper), the line matches will be regarded as an outlier and rejected.

4.2. Free-net BA optimization

Matching of line segments, and consequent identification of line outliers via projective relationships as described in Section 4.1 requires an accurate relative camera estimation. In addition, accurate initialization of cameras and object points increase the likelihood of a rapid convergence to a global minimum during the optimization process described in Section 4.3.

Thus, a free-net BA method, regarded as the golden standard for EOPs, is used to recover relative camera poses and tie points within an arbitrary coordinate system. We set this local coordinate system to the first camera used. First, an incremental BA method, a local optimization solution, is used to reject incorrect point matches and compute the initial values of the

camera for the free-net BA. When adding the last image into the incremental BA model, we consequently update results of the EOPs using a global optimization solution, namely, BA method. In addition, camera distortion is firstly eliminated using a point-based camera self-calibration method, and then calibrated images are processed by the proposed method.

The camera poses optimized by the free-net BA method are termed local EOPs, expressed by (R_L^i, t_L^i) , while the local tie points optimized are represented by X_L^i . Using the locations of GCPs, we can describe the absolute orientation problem to calculate a relative rotation (r), a relative shift (T) and a scale (s) between the local coordinate and GCPs's coordinate. The tie points and the camera locations in the local coordinate can be transformed using

$$\begin{cases} X_G^i = srX_L^i + T \\ t_G^i = srt_L^i + T \end{cases} \quad (18)$$

Here, X_G^i and t_G^i denote tie points and camera translation vectors geo-located in the geodetic coordinate the same as that of the GCP. Similarly, the rotations of initial cameras (R_G^i) in GCPs' coordinate in the proposed method are calculated via

$$R_G^i = R_L^i r^{-1} \quad (19)$$

4.3. HBA optimization using line feature constraints

Using the absolute orientation method described in Section 4.2, we transform the EOPs from a local coordinate system onto the geodetic system describing the GCPs. We can therefore determine the types of 3D line observed by comparing normal vectors of the lines with the Z-axis vector. For a set of corresponding line segments, we calculate the four parameters of the 3D line resulting from the absolute orientation method. The normal vector $n_L = \{n_L^1, n_L^2, n_L^3\}$ of the 3D line is obtained using four calculated parameters in Equation (9), and then its elevation angle can be computed via

$$\theta = \arctan \frac{\sqrt{(n_L^1)^2 + (n_L^2)^2}}{n_L^3} \quad (20)$$

We consider an elevation angle ranging between -5° and $+5^\circ$ as vertical, and impose a corresponding vertical line constraint into the HBA model. Similarly, lines with absolute elevation angles greater than 85° are constrained to be horizontal lines in the HBA model. Lines falling outside these criteria are rejected.

Table 1: The parameters of the tested data, including a simulated data and a real data.

| Parameters | Simulated data | Village data |
|----------------------|-----------------------|-----------------------|
| Tracks | 3 | 3 |
| Area | 100 km \times 10 km | 2 km \times 3.5 km |
| Height (m) | 5,000 | 1,000 |
| Mapping scale | 1:5,000 | 1:1,000 |
| GSD (m) | 0.5 | 0.1 |
| GTPs | 22 | 12 |
| Images | 237 | 90 |
| Images size (pixels) | 7,680 \times 13,824 | 7,680 \times 13,824 |
| 3D points | 1,815 | 6,648 |
| Point projections | 14,242 | 21,734 |
| 3D Lines | 470 | 6,749 |
| Line projections | 2,674 | 22,837 |
| Horizontal lines | 235 | 6,731 |
| Vertical lines | 235 | 18 |

Finally, we input as observations all point features, line segment features, and horizontal and vertical line constraints into the HBA model. An iterative non-linear solution, such as Gauss-Newton, is utilized to estimate the optimal parameters, including 3D tie points, 3D unbounded lines as well as the EOPs in the geodetic coordinate system. The halting criteria of the iterative solution used in this study were either that magnitudes of unknown parameter changes dropped below 10^{-12} per iteration, or that 20 iterations were completed.

5. Experiments and results

5.1. Tested data

In order to verify the feasibility and performance of the proposed method, we have applied the approach to two aerial data sets, one real and one simulated. The overall parameters of the two data are listed in Table 1.

For the simulated data, we have created an aerial scene of 100 km \times 10 km area involving 237 images across 3 tracks. The camera is mounted on an airborne vehicle working at a height of 5,000 m. The mapping scale is 1:5000, and the Ground Sample Distance (GSD) of one pixel is up to 0.5

m. A total of 14,242 image points (1,815 tie points) and 2,674 line segments projections are used for the measured features. Of these, 235 are horizontal lines and 235 are vertical lines. In addition, 22 Ground Truth Points (GTPs) are used to define the global coordinate system and assess the accuracies of sensor orientation methods.

For the real data, 90 aerial images have been captured by a DMC camera, covering $2.0 \text{ km} \times 3.5 \text{ km}$ Shandong province at a mapping scale of 1:1000, located in the north of China. The nominal focal length of the DMC camera is 120 mm, with a pixel pitch of $12 \mu\text{m}$. The images contain significant amounts of human-engineered features (such as buildings and roads). Textures are distributed across 3 tracks, and the GSD reaches up to 0.1 m when the camera has been mounted on an stable Manned Aerial Vehicle (MAV) on the altitude of 1,000 m. 21,734 image points (6,648 3D tie points) and 22,837 line segments projections (6,749 3D lines) are extracted, among which 6,731 lines are horizontal and the 18 are vertical. Additionally, 12 GTPs measured by high-precision GCP receivers and a total station are used to orient the reconstructed model in Beijing-54 coordinate system.

The experimental results of the proposed method were compared with both a standard BA approach in which we used more than 3 GCPs, and one in which only 3 GCPs were used. For simplicity, the standard BA methods using 3, 6 and 14 GCPs are term as “sBA-3”, “sBA-6” and “sBA-14” in this study.

5.2. Simulated results

The geometry of the simulated data can be seen in Figure 6. Red triangles and blue quads denote camera locations and terrain points, respectively.

The distribution of 237 camera locations and 470 3D lines in the XY plane are shown in Figure 7, indicated as blue dots and purple crosses, respectively. 3 GCPs (red triangles) and 21 GTPs (green circles) with no noise are evenly distributed along the 3 tracks. A total of 470 horizontal and vertical lines, which are expressed using purple crosses, are distributed within the tracks. A white Gaussian noise with $\sigma = 0.3$ pixel is added to all image tie points and observed line segments. Note that a cross in Figure 7 contains a horizontal line as well as a vertical line at the same time.

Next, we analyze the accuracy of the camera locations and Check Points (CPs) associated with the proposed method by comparing against two standard BA methods without such line constraints.

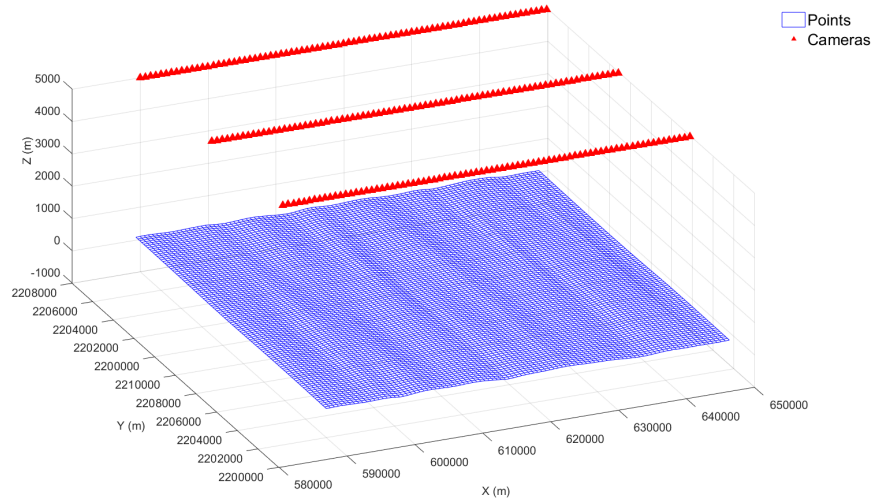


Figure 6: The 3D scene of the simulated data.

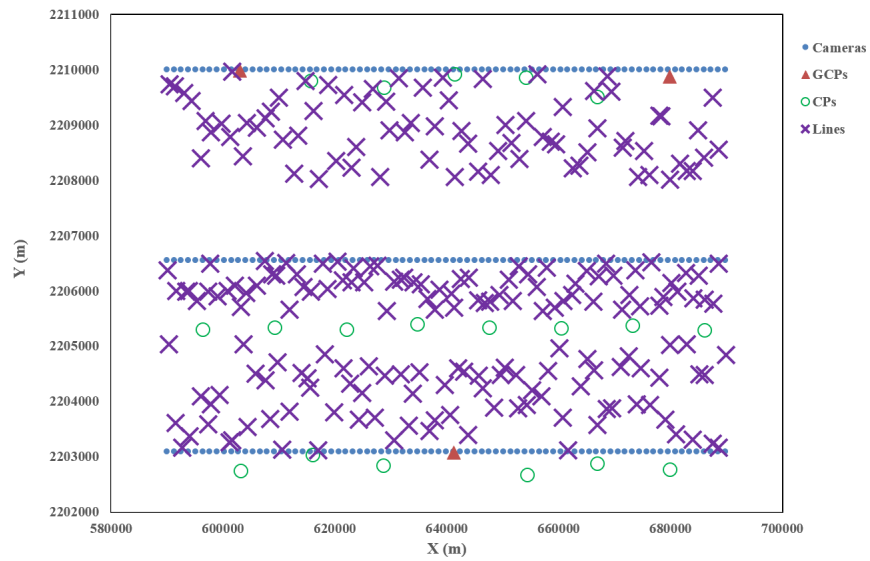


Figure 7: Simulated Camera Locations, GCPs, CPs and Lines, as distributed on the XY plane.

The residuals of the camera locations as estimated by the three methods are shown in Figure 8. The results of our method, using 3 GCPs and line constraints is depicted by a red curve, whilst the results of a standard BA using both 14 GCPs (sBA-14) and 3 GCPs (sBA-3) without line constraints, are shown using blue and green curves respectively. Note that the 14 GCPs used in the standard BA method are distributed on the first and third tracks, with the remaining points on the second track set to CPs.

From the residuals plot, it would appear that sBA-14 provides a superior solution to that of both sBA-3 and the proposed method (as might be expected, given all the additional information available) but our solution does significantly outperform that of sBA-3.

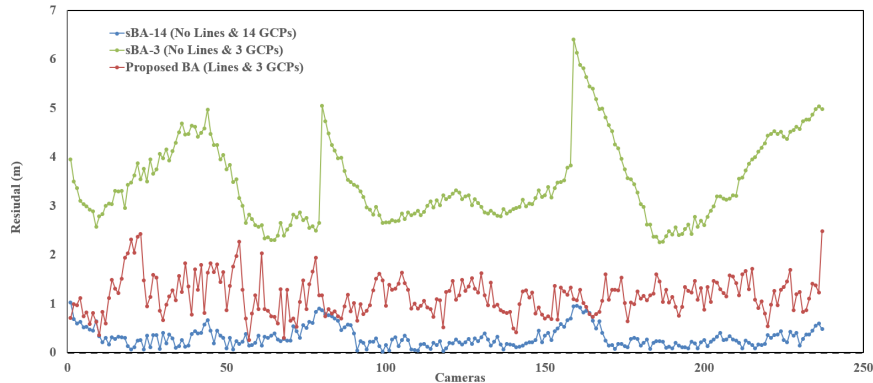


Figure 8: The residuals of camera locations are recovered by two no-lines-based BA methods using 14 and only 3 GCPs, as well as the proposed method.

The Root-Mean-Square-Error (RMSE) of the CPs and camera locations are shown in Figure 9. For the sBA-14 method, the least RMSEs of 0.245 m and 0.372 m can be obtained when assessing the accuracy of the CPs and the camera locations. The accuracy of the CPs and the camera locations with 0.603 m and 1.22 m can be obtained by the proposed method, better than the accuracy of 2.312 m and 3.566 m associated with the BA method using only 3 GCPs and no lines.

Generally speaking, by applying line constraints on the simulated data, we see an improvement in accuracy of approximately 70%.

5.3. Village data results

Real world performance was assessed using our “Village” dataset. The scene is of a built-up rural environment, rich with engineered structures ex-

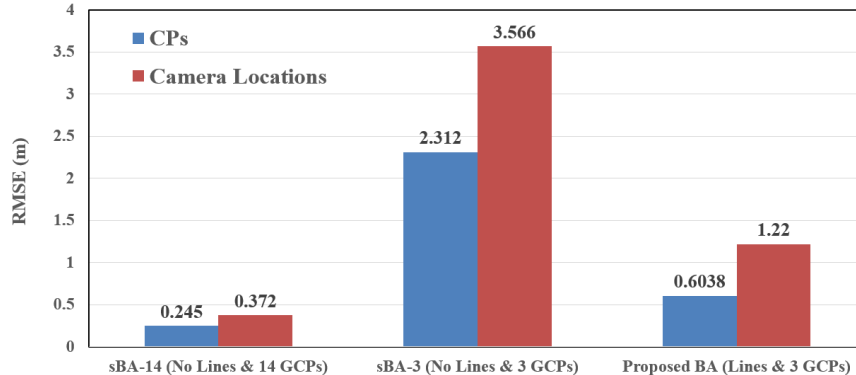


Figure 9: RMSEs of CPs and Camera locations recovered by three BA methods.

hibiting the distinct line features to which our HBA approach is suited. Using the detection and matching algorithm described in Section 4, we extracted a total of 6,749 3D lines from 90 images. By means of example, we show the 14th image together with extracted line segments in Figure 10. Each line segment is represented using a red line together with an index number. We can see that the majority of extracted segments are horizontal, as might be expected given the observation geometry. In addition, 5 vertical edges of the building (668th, 678th, 679th, 685th, and 686th lines) are also detected and contribute horizontal constraints to the HBA model.

The camera locations, GCPs and CPs of the Village data in XY direction are depicted in Figure 11, in which blue dots, red triangles and green circles indicate the cameras, GCPs and CPs, respectively. For the sBA-3 method and the proposed method, 3 red GTPs serve as the GCPs, used in the BA, and the remaining 9 GTPs are used as CPs to check the accuracy of the BA model. In the figure, CPs are indexed and marked with green triangles. For the sBA-6 method, the first three CPs (1st, 2nd and 3th) are treated as GCPs together with the points indicated by red triangles, with the remaining 6 points (4th ~ 9th) used for the assessment of the estimation accuracy.

The extracted 21,734 point projections and 22,837 line projections form our observations, providing our HBA with 118,746 constraints, as well as 47,480 unknown parameters. The plane and height residuals of the CPs estimated by the three methods are shown in Figure 12, in which the blue, green and red curves describe the results of the sBA-6, sBA-3, and the proposed method respectively. Comparing the last 6 CPs, we can see that the

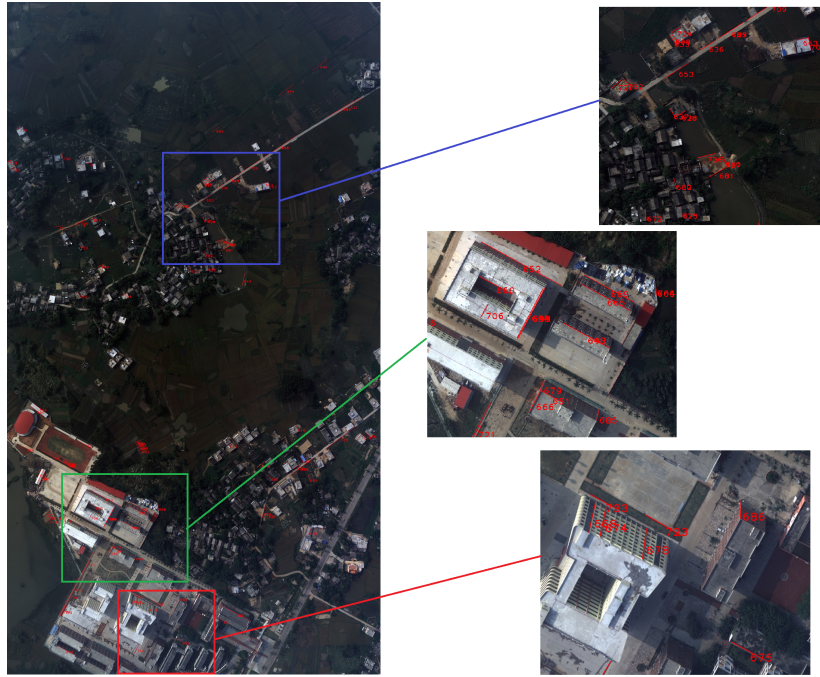


Figure 10: The extracted line from the 14th image serve as the constraints of the proposed method.

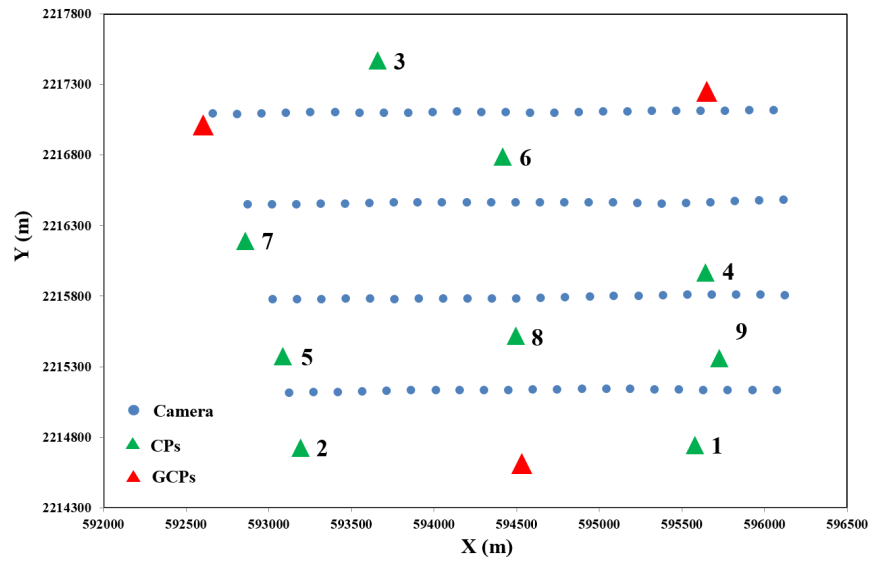


Figure 11: Camera locations, GCPs and CPs of the Village data, as distributed on the XY plane.

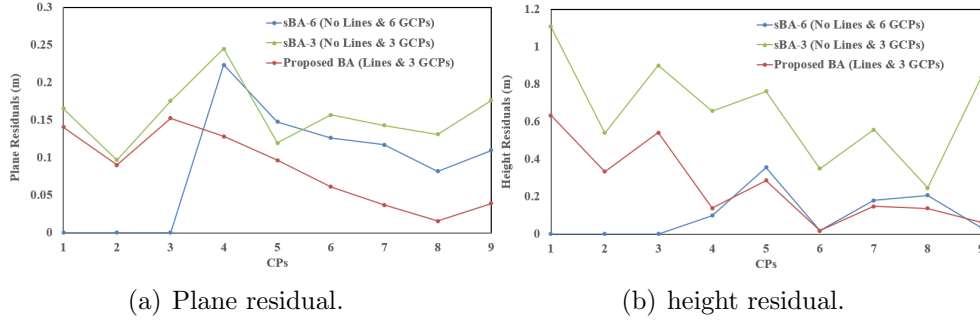


Figure 12: RMSEs of CPs are evaluated by two traditional methods and the proposed method.

Table 2: The RMSEs of the CPs of the Village dataset resulting from sBA-6, sBA-3 and the proposed method.

| Residual (m) | | sBA-6 (No Lines & 6 GCPs) | sBA-3 (No Lines & 3 GCPs) | Proposed BA (Lines & 3 GCPs) |
|--------------|--------|---------------------------|---------------------------|------------------------------|
| Max | East | 0.1474 | 0.1358 | 0.0874 |
| | North | 0.2026 | 0.2062 | 0.0897 |
| | Height | 0.3545 | 1.1073 | 0.6333 |
| RMSE | East | 0.0764 | 0.0887 | 0.0400 |
| | North | 0.1190 | 0.1348 | 0.0876 |
| | Height | 0.1877 | 0.7093 | 0.3248 |
| | Plane | 0.1414 | 0.1614 | 0.1191 |
| | Total | 0.2350 | 0.7274 | 0.3388 |

three methods share a similar plane residual, but that the Z accuracy of the proposed method is better than that of sBA-3 when the same GCPs are integrated into the BA model.

In addition, we calculate the statistics of RMSEs of the CPs of the Village data estimated by the three methods in Table 2. With six high-quality GCPs, sBA-6 without line constraints provides RMSE accuracy of 0.235 m. However, removing 3 GCPs increases the error to 0.7274 m. Using our HBA on the same 3 GCPs improves the accuracy to 0.3355 m. An improvement of the accuracy is again, approximately 50%.

We plot 3D point clouds associated with three methods in Figure 13, in which green, blue and red dots represent the results of sBA-6, sBA-3 and the proposed method respectively. To clearly distinguish the differences, we randomly pick up 10 points in the three arbitrary regions, labeled in black

boxes, and calculate the residuals of sBA-3 and the proposed method when comparing against sBA-6. In the three subfigures, the blue curves represents the residuals of sBA-3, while the red curves are those of the proposed method. We see that the average residuals of points in sBA-3 are 0.6 m, but those of the proposed method are only 0.2 m. Thus, we can conclude that adding horizontal and vertical line constraints can control the error propagation, providing the similar function of another 3 GCPs used in sBA-6.

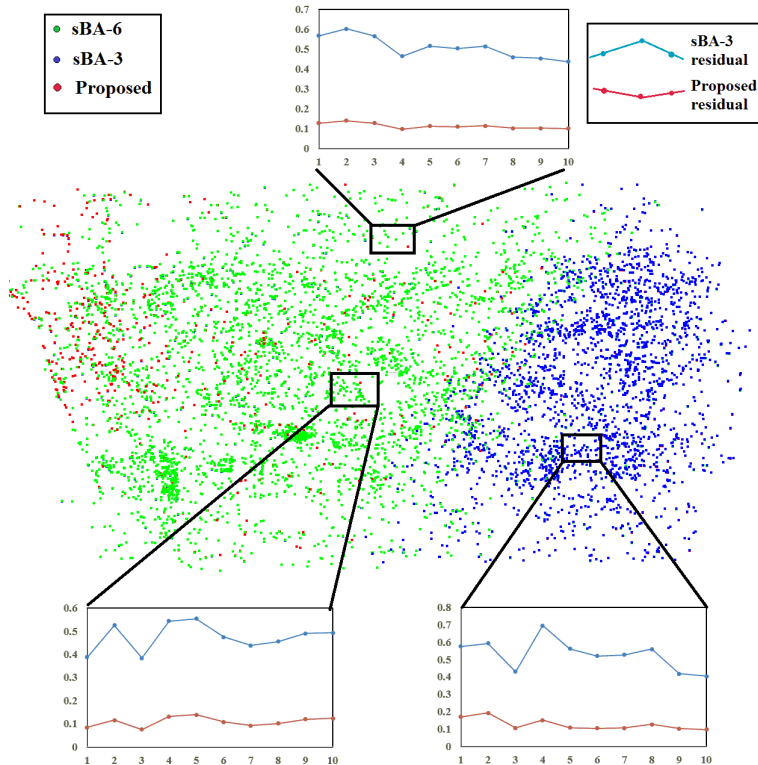


Figure 13: 3D points are obtained by two standard methods and the proposed method.

At the same time, we also analyze the accuracies of the estimated camera locations of the proposed method by comparing with sBA-3. The Z values of the 90 camera locations estimated by three methods are shown in Figure 14, in which three subfigures are the Z values of cameras 1-10, 41-50 and 61-70. It demonstrates that the accuracy of the proposed method is closer to that of sBA-6, and better than those of sBA-3.

Residuals of point estimates on the plane for each of sBA-3 and the pro-

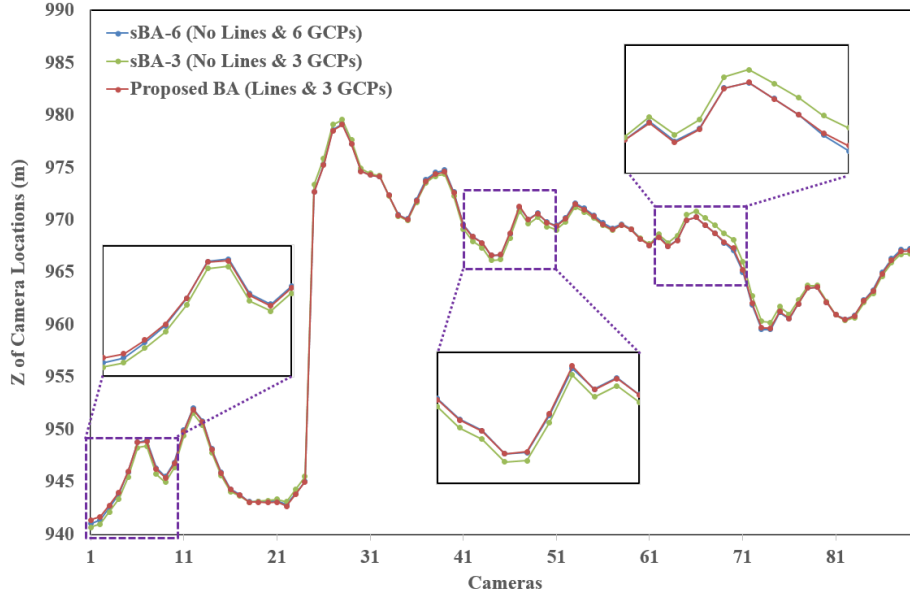


Figure 14: The heights of each cameras are estimated by two standard methods and the proposed method.

posed BA compared to those of sBA-6 are shown in Figure 15. The blue curve is shows sBA-3, while the red one is the proposed method. It is clear that the camera locations estimates on the plane estimated by the proposed method are also closer to those of the sBA-6.

In summary, from the experimental results of the simulated and real data, we can see that utilization of line constraints can significantly improve the accuracy of our sensor orientation estimates. Whilst it remains preferable to simply introduce more GCPs into a survey where possible, this is not always practicable. In such cases, our method can provide auxiliary observations and control information to reduce error propagation. Thus, the accuracy of camera orientation estimates can be improved.

6. Discussion

In this section, we principally discuss the impact different line constraint and weighting ratio parameters have on estimate performance, using both the simulated data and real data in Section 5.

For line constraints, we mainly focus on any accuracy improvement observed between general line constraints and horizontal/vertical line constraints.

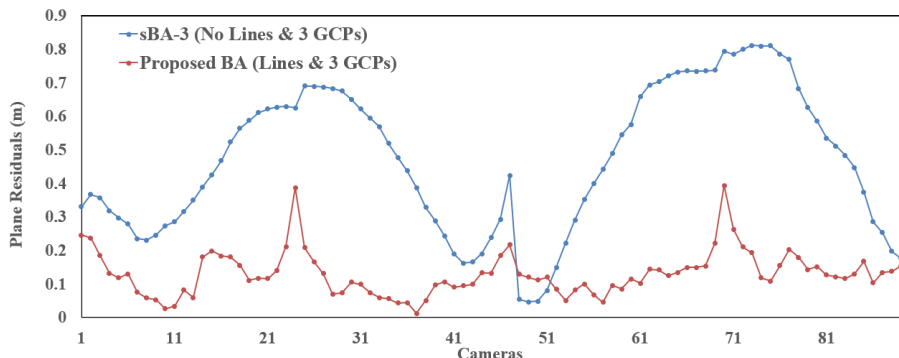


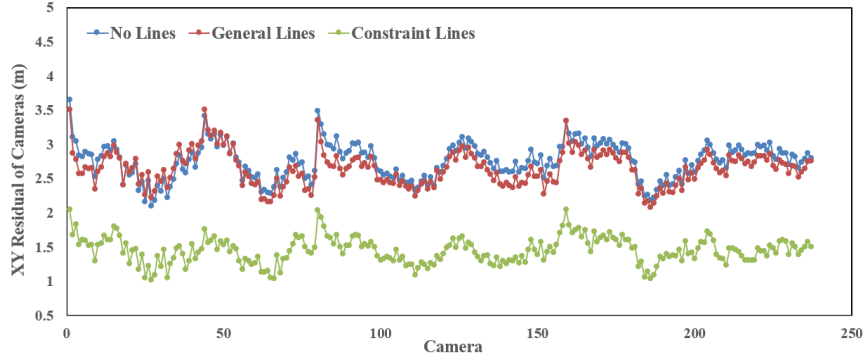
Figure 15: The plane residuals of the estimated cameras by calculating the difference with the estimated plane camera locations of sBA-6.

For the weighting ratio parameter λ , different geometric meanings of the four components that form the objective function make it difficult to precisely express the uncertainties of the observations in form of mathematical equation. Thus, we determine an appropriate parameter for use in Section 5 empirically below.

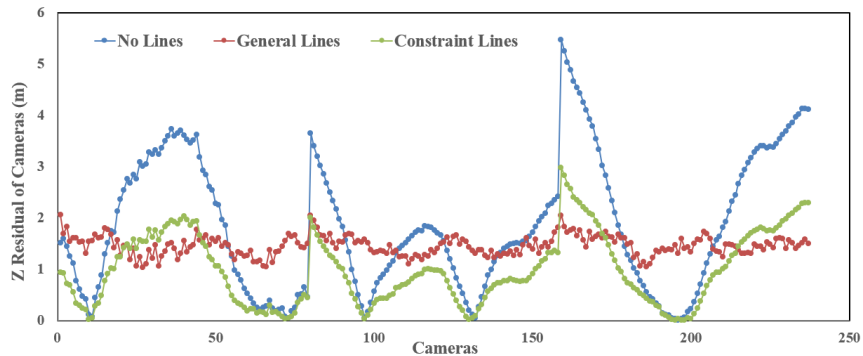
Firstly, we show 3 BA results, the first one (“No Lines”) is a BA using only image point observations. The second, (“General Lines”) is a BA using both image point and general line segment observations, where the orientation of the lines are unconstrained. The third is the proposed method representing a BA with horizontal and vertical line constraints, labelled “Constraint Lines”.

The residuals of the camera locations in the simulated data and the real data are shown in Figure 16 and Figure 17, respectively. In addition, RMSEs are depicted in Figure 18. From these above 3 figures, it is clearly seen that the accuracy of point features can be slightly improved by adding general lines, however much more significant improvement can be seen by adding horizontal and vertical line constraints.

We posit that horizontal and vertical lines provide more constraints in plane and height directions, which reduces the large scale error propagations of point/line observation, by bounding any error to a small locality. We suspect that the only-modest improvements in accuracy of the general line approach over a point only approach is that, although 2D line segments can provide more information (constraint), the poorer accuracy of the end-points obtained by 2D line segment extraction results in less accurate observations. However, the added horizontal and vertical line constraints can significantly



(a) XY-direction residuals.

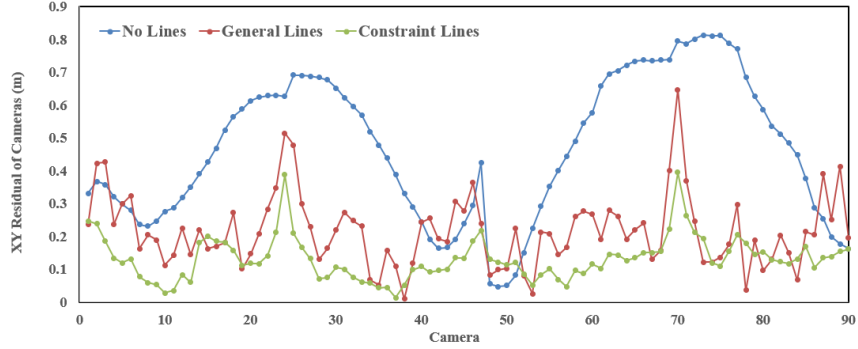


(b) Z-direction residuals.

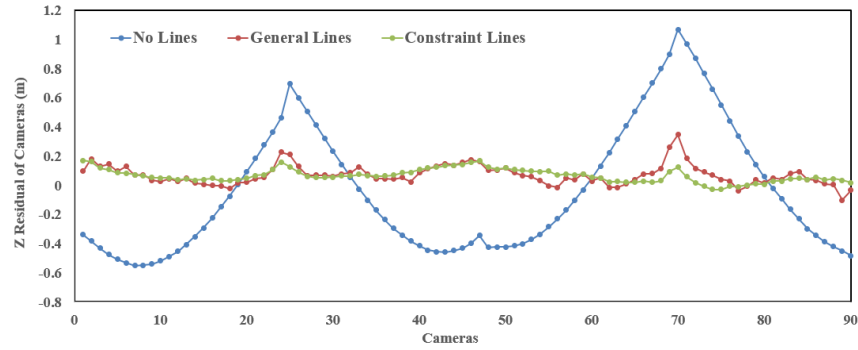
Figure 16: Residuals of 237 simulated camera locations recovered by adding different line constraints.

reduce the impact this has on the wider adjustment, by bounding any error to a small locality. Thus, in our method, horizontal and vertical lines are included as line observations.

Secondly, the weighting ratio parameter, namely horizontal/vertical line constraints w.r.t. point/line constraint, affecting the weighting matrices used in the least squares problem are defined in Equation (15). In this study, the parameter was manually set, and their suggested values are discussed in the experimental analyses as follows. The simulated and real dataset were used to discuss the accuracy variation with the parameter. Here, we set the parameter λ as 10^{-4} , 10^{-2} , 10^0 , 10^2 , 10^3 , 10^4 , 10^5 and 10^6 . The curves of camera locations residuals of the simulated data and the real data are shown in Figure 19 and Figure 20, while RMSE is also shown in Figure 21. We



(a) XY-direction residuals.



(b) Z-direction residuals.

Figure 17: Residuals of 90 Village camera locations recovered by adding different line constraints.

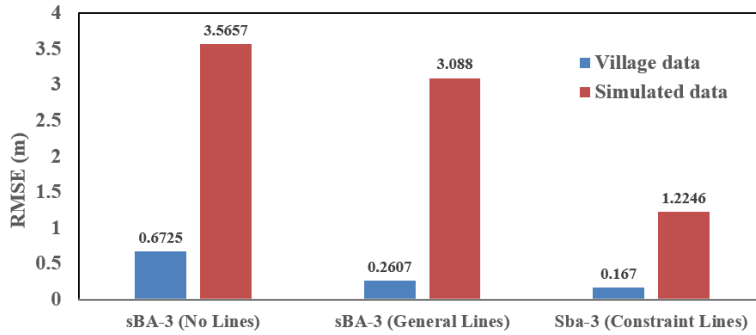


Figure 18: RMSE of camera locations of two datasets recovered by adding different line constraints.

discover that, for both two datasets, the best accuracy can be obtained using the parameter 10^3 . When the parameter is larger than 10^3 or smaller than 10^2 , poorer accuracy is estimated. Thus, 10^3 is used in the experiments of this study. We appreciate that this empirical approach is less desirable than development of a means of relating the underlying units of covariance, but leave this for further work.

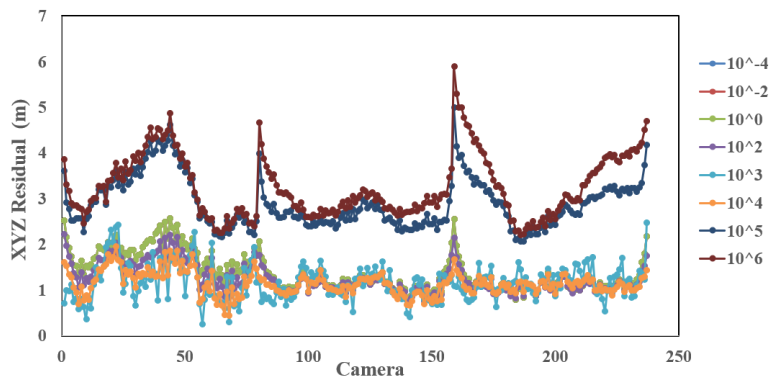


Figure 19: Curves of camera locations residuals of simulated data obtained when different weighting ratio parameters are inputted.

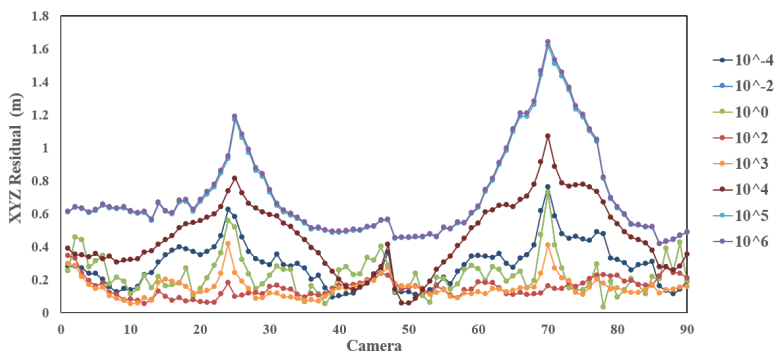


Figure 20: Curves of camera locations residuals of Village data obtained within different weighting ratio parameters.

7. Conclusions and Future work

The potential for increasing the accuracy of indirect sensor orientation estimates has already been verified in the state-of-the-art work of (Gerke,

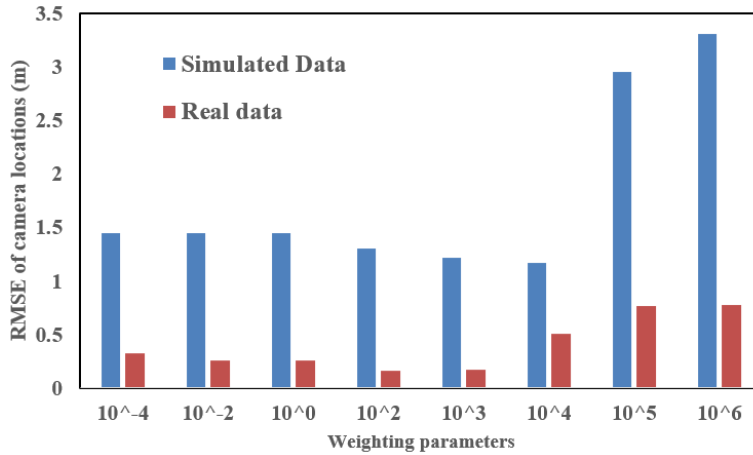


Figure 21: RMSE of camera locations of two datasets calculated within different weighting ratio parameters.

2011), who showed horizontal and vertical building features could successfully be used as additional constraints. However, some manual work was still involved to define the constraints, since end-points of structure edges had to be manually measured, with the horizontal and vertical constraints defined based on geometric relationship between the two end-points. This study proposed a new BA method without manual work, in which unbounded line features replace end-points to define the constraints.

Our proposed hybrid BA method contains two types of observed features, namely, image points and 2D line segments. A 3D line is parameterized using normal vectors of two planes, and then normal vectors of the plane passing through other corresponding line segments are calculated based on triple tensor model. In the cost functions of our BA method, image point projections, line segment projections, horizontal normal vectors and vertical ones are jointly used for the non-linear optimization observations to estimate camera poses, 3D points and 3D lines.

Using both simulated data and real data, we found that a 50% increase in accuracy of the sensor orientation can be obtained by the proposed method, when comparing with a standard BA method (without line feature constraints). In addition, by performing a BA with arbitrarily-oriented line features as well as line-features that were constrained to the horizontal and vertical, we determined that it was the horizontal and vertical constraints, and not the additional observations introduced by use of line-features, that

provided the increase in estimate accuracy.

In this work we have used pre-rectified images as a source for our bundle adjustment, and thus not included camera distortion parameters as unknowns. There are many real-world applications where knowledge of the intrinsic parameters will not be known a-priori. Therefore our future work will seek to introduce and test the robustness of this approach with the inclusion of camera calibration estimation.

Further, notions of the position uncertainty of 2D and 3D points, and the mechanisms for introducing them into a bundle adjustment, are well established in our community. Characterising error mechanisms and uncertainties surrounding line features when used as constraints in an adjustment is something not well explored. We feel fruitful research in this area might be made, with the ultimate aim to parameterize line feature uncertainty's such that they can be meaningfully included in our adjustment approach.

Acknowledgment

The authors would like to thank Xing Tian Di (Beijing) Co., Ltd for providing a set of aerial images and auxiliary data used for experimental assessment.

References

- Alamus R., Kornus W., Talaya J., 2006. Studies on DMC geometry. *ISPRS Journal of Photogrammetry and Remote Sensing*, 60(6), 375–386.
- Gneeniss A S., Mills J P., Miller P E., 2013 Reference LiDAR surfaces for enhanced aerial triangulation and camera calibration. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, 1, 111–116.
- Fan B., Wu F., Hu Z., 2012. Robust line matching through line-point invariants. *Pattern Recognition*, 45(2), 794–805.
- Gerke M., 2011. Using horizontal and vertical building structure to constrain indirect sensor orientation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3), 307–316.
- Gerke M., 2009. Dense matching in high resolution oblique airborne images. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, 38: W4.

- Gerke M., Kerle N., 2011. Automatic structural seismic damage assessment with airborne oblique Pictometry imagery. *Photogrammetric Engineering & Remote Sensing*, 77(9), 885–898.
- Gomez-Ojeda, R., Zuniga-Noel, D., Moreno, F. A., Scaramuzza, D., Gonzalez-Jimenez, J. 2017. PL-SLAM: a Stereo SLAM System through the Combination of Points and Line Segments. arXiv preprint arXiv:1705.09479.
- Gülch E., 2012. Photogrammetric measurements in fixed wing UAV imagery. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Melbourne, Australia, August, 381-386.
- Hartley R., Zisserman A., 2003. *Multiple view geometry in computer vision*. Cambridge university press.
- Habib A F., Morgan M., Lee Y R., 2002. Bundle adjustment with self-calibration using straight lines. *The Photogrammetric Record*, 17(100), 635–650.
- Hofer M., Maurer M., Bischof H. 2017. Efficient 3D scene abstraction using line segments. *CVIU* 157(April):167-178.
- James M R., Robson S., 2012. Straightforward reconstruction of 3D surfaces and topography with a camera: Accuracy and geoscience application. *Journal of Geophysical Research: Earth Surface*, 117(F3).
- Jeong, W. Y., Lee, K. M. 2006. Visual SLAM with line and corner features. *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2570–2575.
- Lu X., Yao J., Li K., Li L., 2015. CannyLines: A parameter-free line segment detector. *IEEE International Conference on Image Processing (ICIP)*, 507–511.
- Ramalingam, S., Bouaziz, S., Sturm, P. 2011. Pose estimation using both points and lines for geo-localization. *IEEE International Conference on Robotics and Automation (ICRA)*, 4716–4723.
- Roberts K S., 1988. A new representation for a line. *Computer Society Conference on Computer Vision and Pattern Recognition*, 635–640.

- Teo T., Chen L., Liu C., Tung Y., Wu W., 2010. DEM-aided block adjustment for satellite images with weak convergence geometry. *IEEE Transactions on Geoscience and Remote Sensing*, 48(4), 1907–1918.
- Schenk T., 2004. From point-based to feature-based aerial triangulation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58(5), 315–329.
- Sun, Y., Zhao, L., Huang, S., Yan, L., Dissanayake, G., 2014. L2-SIFT: SIFT feature extraction and matching for large images in large-scale aerial photogrammetry. *ISPRS Journal of Photogrammetry and Remote Sensing*, 91, 1–16.
- Sun Y., Sun H., Yan L., et al., 2016. RBA: Reduced Bundle Adjustment for oblique aerial photogrammetry. *ISPRS Journal of Photogrammetry and Remote Sensing*, 121, 128–142.
- Vakhitov, A., Funke, J., Moreno-Noguer, F. 2016. Accurate and linear time pose estimation from points and lines. *European Conference on Computer Vision (ECCV)*, 583-599.
- Yastikli N., Jacobsen K., 2005. Influence of system calibration on direct sensor orientation. *Photogrammetric Engineering & Remote Sensing*, 71(5): 629–633.
- Yuan X., Fu J., Sun H., Toth C, 2009. The application of GPS precise point positioning technology in aerial triangulation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(6), 541–550.
- Yuan X., Chou X., Wang S., 2004. Image orientation by combined bundle adjustment with fixed imageries. *International Archives for Photogrammetry and Remote Sensing*, 35(B2), 34–38.
- Zhang Y., Hu B., Zhang J., 2011. Relative orientation based on multi-features. *ISPRS journal of photogrammetry and remote sensing*, 66(5), 700–707.
- Zhang Y., Hu K., Huang R., 2012. Bundle adjustment with additional constraints applied to imagery of the Dunhuang wall paintings. *ISPRS journal of photogrammetry and remote sensing*, 72, 113–120.

- Zhao L., Huang S., Yan L., Dissanayake G., 2015. A new feature parametrization for monocular SLAM using line features. *Robotica*, 33(3), 513-536.
- Zhao L., Huang S., Sun Y., Yan L., Dissanayake, G. 2015. ParallaxBA: Bundle adjustment using parallax angle feature parametrization. *International Journal of Robotics Research (IJRR)*, 34(4-5), 493-516.
- Zheng, F., Tsai, G., Zhang, Z., Liu, S., Chu, C. C., Hu, H. 2018. PI-VIO: Robust and Efficient Stereo Visual Inertial Odometry using Points and Lines. arXiv preprint arXiv:1803.02403.
- Zuo X., Xie X., Liu Y., Huang G. Robust visual SLAM with point and line features. 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). 2017