remote sensing

MDPI

Article

Optimal Self-Calibration Strategies in the Combined Bundle Adjustment of Aerial–Terrestrial Integrated Images

Linfu Xie [1,2], Han Hu [3], Qing Zhu [3], Xiaoming Li [1], Xiang Ye [1], Renzhong Guo [1], Yeting Zhang [2], Xiaoqiong Qin [4] and Weixi Wang [1,*]

1 Research Institute for Smart Cities, School of Architecture and Urban Planning, Shenzhen University & Guangdong Key Laboratory of Urban Informatics & Shenzhen Key Laboratory of Spatial Smart Sensing and Services, Shenzhen 518060, China; linfuxie@szu.edu.cn (L.X.); lixming@szu.edu.cn (X.L.); yexiang@szu.edu.cn (X.Y.); guorz@szu.edu.cn (R.G.)

2 State Key Laboratory of Information Engineering in Surveying Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China; zhangyeting@263.net

3 Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu 611756, China; han.hu@swjtu.edu.cn (H.H.); zhuq66@263.net (Q.Z.)

4 College of Civil and Transportation Engineering, Shenzhen University, Shenzhen 518060, China; xqqin@szu.edu.cn

* Correspondence: wangwx@szu.edu.cn; Tel.: +86-135-9031-7811

Abstract: Accurate combined bundle adjustment (BA) is a fundamental step for the integration of aerial and terrestrial images captured from complementary platforms. In traditional photogrammetry pipelines, self-calibrated bundle adjustment (SCBA) improves the BA accuracy by simultaneously refining the interior orientation parameters (IOPs), including lens distortion parameters, and the exterior orientation parameters (EOPs). Aerial and terrestrial images separately processed through SCBA need to be fused using BA. Thus, the IOPs in the aerial–terrestrial BA must be properly treated. On one hand, the IOPs in one flight should be identical for the same images in physics. On the other hand, the IOP adjustment in the cross-platform-combined BA may mathematically improve the aerial–terrestrial image co-registration degree in 3D space. In this paper, the impacts of self-calibration strategies in combined BA of aerial and terrestrial image blocks on the co-registration accuracy were investigated. To answer this question, aerial and terrestrial images captured from seven study areas were tested under four aerial–terrestrial BA scenarios: the IOPs for both aerial and terrestrial images were fixed; the IOPs for only aerial images were fixed; the IOPs for only terrestrial images were fixed; the IOPs for both images were adjusted. The cross-platform co-registration accuracy for the BA was evaluated according to independent checkpoints that were visible on the two platforms. The experimental results revealed that the recovered IOPs of aerial images should be fixed during the BA. However, when the tie points of the terrestrial images are comprehensively distributed in the image space and the aerial image networks are sufficiently stable, refining the IOPs of the terrestrial cameras during the BA may improve the co-registration accuracy. Otherwise, fixing the IOPs is the best solution.

Keywords: image orientation; UAVs; photogrammetry; data integration

check for updates

Citation: Xie, L.; Hu, H.; Zhu, Q.; Li, X.; Ye, X.; Guo, R.; Zhang, Y.; Qin, X.; Wang, W. Optimal Self-Calibration Strategies in the Combined Bundle Adjustment of Aerial–Terrestrial Integrated Images. *Remote Sens.* **2022**, *14*, 1969. https://doi.org/10.3390/rs14091969

Academic Editor: Biswajeet Pradhan

Received: 11 March 2022
Accepted: 17 April 2022
Published: 19 April 2022

1. Introduction

Multi-view images captured by aerial or unmanned aerial vehicle (UAV) platforms have become a major source of data in 3D city modeling projects [1–4]. To alleviate occlusions and increase observation redundancy, in the last decade, images captured by different cameras, such as vertical and oblique views in multi-camera systems, are combined to product photo-realistic 3D models with better geometry quality and textures [5–7].

Because the accuracy of modeling image distortions and orientations directly affects the product quality in subsequent image-processing steps such as dense image matching, 3D mesh generation, and texture mapping, numerous methods have been developed to recover EOPs and IOPs (including lens distortion parameters) [8–12].

Among existing photogrammetry research and engineering practices, self-calibrated bundle adjustment (SCBA), through which IOPs and EOPs are simultaneously estimated according to image tie points, is an effective method for decreasing re-projection errors in 2D space and intersection errors in 3D space [13–17]. Different from the traditional laboratory or field calibration processes, self-calibration (SC) methods treat IOP calibration as part of routine photogrammetric procedures in every project through bundle adjustment (BA), whether the cameras have been pre-calibrated or not [18].

In recent years, terrestrial images captured by hand-held cameras or mobile mapping platforms have been integrated with aerial views through structure-from-motion and multi-view stereo pipelines to produce better 3D maps and models [19–22]. Due to large differences in viewpoint and scale and possible illumination conditions, automatic feature matching for cross-platform images is non-trivial work [21]. The numbers and distributions of cross-platform tie points are not as favorable as those for inner-platform images. Hence, in 3D modeling applications that integrate images captured by aerial and terrestrial platforms, images taken in the same platform are often first aligned through BA using only inner-platform tie points. Then, aerial and terrestrial images are co-registered using a cross-platform involving BA [19–21]. Although IOPs should be refined during the inner-platform BA, it remains unclear whether the IOPs in the cross-platform BA should be fixed.

On one hand, the IOPs are recovered through SCBA with inner-platform tie points, and the images used in the cross-platform remain unchanged; thus, the IOPs in the second BA should be physically the same as those in the first BA. IOP fixation could reduce the number of unknown parameters and stabilize the calculation of the nonlinear least square problem. On the other hand, according to SCBA theory, refining the IOPs in the cross-platform BA may mathematically improve the modeling quality of the image formatting process and thus enhance the co-registration quality between the aerial and terrestrial images.

Both strategies seem reasonable. Hence, to investigate the optimal SC strategy for the BA of aerial–terrestrial integrated images, four aerial–terrestrial BA settings were experimentally compared and analyzed in this study. According to the experimental results, recommendations on the integration of aerial and terrestrial images blocks in BA are provided.

The remainder of this paper is organized as follows: Section 2 reviews the existing work on SCBA. Section 3 introduces the four plausible SCBA strategies and the experimental datasets and procedure. Section 4 reports the experimental results and analysis. Section 5 presents the discussion, conclusions, and future perspectives.

2. Related Works

The geometry quality of image-based 3D mapping products largely relies on the precision of the recovered image IOPs and EOPs [23]. In traditional photogrammetry engineering, the IOPs of metric cameras are first calibrated in the laboratory or field before image capture [9,13,24]. After image collection, EOPs are recovered through BA according to image correspondence and a few ground control points [4,25,26]. Previous investigations have proved that when a sufficiently accurate camera model is used, the 3D mapping inaccuracy related to systematic errors in IOPs is negligible [27]; however, this is not the case in close-range photogrammetry [28,29].

With the rapid development of UAVs, consumer-level cameras are widely used in small- or clustered-area survey tasks [30–35]. Compared with traditional aerial photogrammetry, which collects images with only vertical views, adding oblique cameras could not only result in better façade information but also favor geometry measuring accuracy due to the larger intersection angles between overlapped images [6,36,37]. This advantage is more noticeable for UAV photogrammetry since the flight plans are far more flexible [38,39]. While image collection is relatively easy, a rigorous laboratory or field camera calibration process is often neglected. Moreover, the images are often captured by unprofessional operators with inferior geometry networks. Furthermore, the sensor stability may be imperfect, as 3D mapping and modeling are not the main purposes of the camera design.

Hence, in most UAV photogrammetry applications, SCBA is commonly adopted to refine the IOPs to improve the 3D accuracy of object mapping [3,26,32,40]. In standard SCBA, the IOPs are treated as unknown parameters with initial observations; moreover, the IOPs, EOPs, and 3D coordinates of tie points during the BA process are refined according to the collinearity equation [18]. Regarding the sensor stability under different temperatures and humidities, SCBA is often conducted in every image block, because the images are collected under various conditions.

Apart from the conventional focal length and the location of principal points, additional parameters are used to describe the distortions that occur between 3D points and their locations in 2D images, because the image formatting process is not a perfect perspective transformation [24]. As pointed out in previous works [41], there are two major categories of additional parameters: the physical and mathematical models [13,41,42]. The physical models simulate the systematic errors caused by optics, while the mathematical models approximate the simulation process through algebra expansion.

The Brown model [13] is the most widely used model in close-range photogrammetry, and it has been incorporated into numerous image-processing packages and commercial softwares in United States, China, and Russia [43–45]. Although SCBA implementation with additional parameters may improve geometric accuracy in practice, the correlation between parameters may weaken the BA process [18]. Moreover, to obtain satisfactory results, a large number and good distribution of image tie points are required [46].

In recent years, 3D environmental modeling applications have combined images captured by airborne cameras and terrestrial platforms [20,21,39]. Images captured by a singular platform and processed through BA also require cross-platform BA for accurate co-registration in 3D mapping and modeling tasks [19].

In addition to the viewing perspective and image scales, images obtained by aerial and terrestrial platforms also vary in other aspects. First, the networks for aerial images are often more stable, because aerial views have more image connectivity than terrestrial views. Moreover, owing to the differences in looking directions and scenes, terrestrial images might have weak textures at the corners or edges. Thus, the distributions of automatic tie points are often more unsymmetrical than for aerial datasets.

However, it is still unclear whether it is best to re-compute the IOPs during the BA of aerial and terrestrial images. On one hand, the SCBA are already adopted in the BA, which align images captured by the same platform; the IOPs can be treated as stable because the images remain unchanged [19]. Moreover, IOP fixation can reduce the number of unknown parameters, which may stabilize the EOP calculation process and possibly enhance the estimation accuracy. Furthermore, the additional cross-platform tie points that align aerial and terrestrial images can result in uneven tie point distribution [39]. Because more tie points (mainly the cross-platform tie points) are incorporated into the BA process, the image geometric network is varied. From a mathematical viewpoint, the integrated refinement of both IOPs and EOPs may improve the recovery of the image formatting process and thereby improve the co-registration between aerial and terrestrial images.

To investigate the optimal SC strategy for the BA of aerial–terrestrial integrated images after the inner-platform SCBA, four SC strategies for the BA were compared and analyzed using real datasets.

3. Experimental Settings and Datasets

3.1. Experimental Hypothesis and Settings

To evaluate the performances of different SCBA strategies for aerial–terrestrial image blocks (Figure 1), the following procedures were adopted. Starting from aerial and terrestrial images with initial EOPs and IOPs in seven study sites, a SCBA was separately adopted for aerial and terrestrial images, and the updated EOPs and IOPs were obtained. Then, BA was implemented to integrate aerial and terrestrial images. In the BA, four SC strategies were used: (1) fixing the IOPs for both aerial and terrestrial images (FA_FT); (2) fixing the IOPs for aerial images while refining the IOPs for terrestrial images (FA_RT); (3) refining the IOPs for aerial images while fixing the IOPs for terrestrial images (RA_FT); (4) refining the IOPs for both aerial and terrestrial images (RA_RT). After the implementation of BA, IOPs and EOPs for aerial and terrestrial images were obtained.

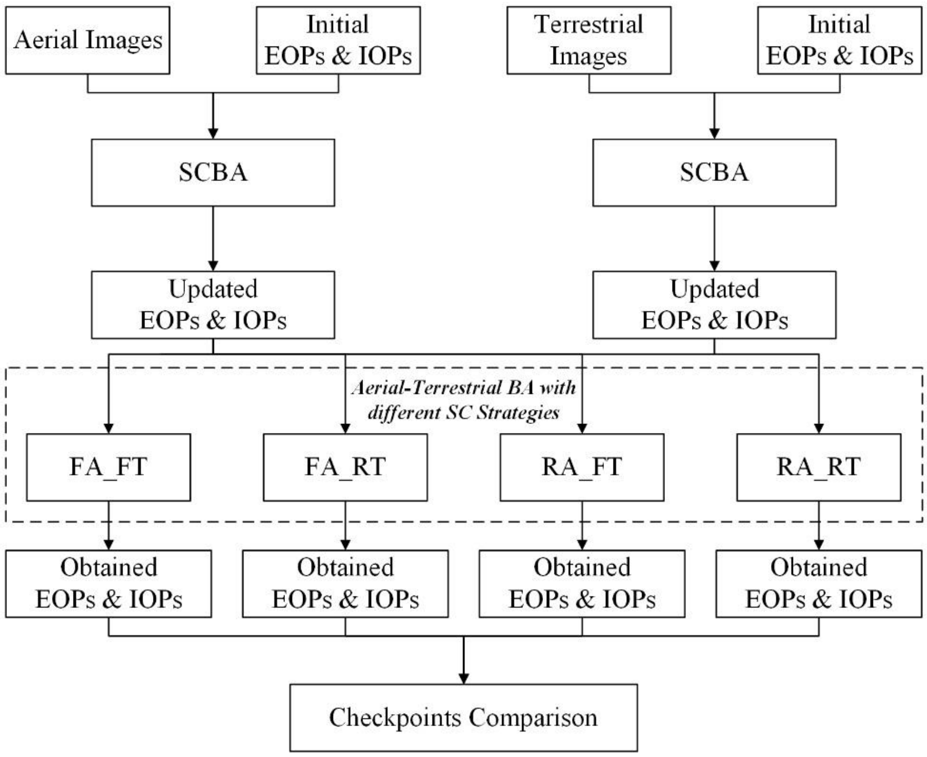

Figure 1. The overall workflow of the study.

In this study, the camera was the classic pinhole model; the lens distortion parameter is given in Equation (1).

$$x = F \cdot D(\prod R(X - X_c)) + x_0$$
$$F = \begin{bmatrix} f & 0 \\ 0 & f \end{bmatrix} \quad (1)$$

$$\prod(u, v, w) = (u/w, v/w) \quad (2)$$

where X, a 3D column vector, represents the 3D position of a point in the object space, and x, a 2D vector, denotes the position of the corresponding point in the image space. X_c and R are the EOPs, where X_c is the 3D position of the camera center, while R represents a 3×3 rotation matrix that maps between the axes of world coordinates and the camera axes. f denotes the focal length, and x_0 denotes the position of the principal point. $\prod : R^3 \to R^2$ represents the perspective projection defined by Equation (2), and u, v, and w represent the coordinates of a 3D point in the camera space. The additional parameters of the lens distortion are given by $D : R^2 \to R^2$, where

$$D(u, v) = \begin{pmatrix} (1 + k_1 \cdot r^2 + k_2 \cdot r^4 + k_3 \cdot r^6) \cdot u + 2p_2 \cdot u \cdot v + p_1 \cdot (r^2 + 2u^2) \\ (1 + k_1 \cdot r^2 + k_2 \cdot r^4 + k_3 \cdot r^6) \cdot v + 2p_1 \cdot u \cdot v + p_2 \cdot (r^2 + 2v^2) \end{pmatrix} \quad (3)$$
$$r^2 = u^2 + v^2$$

Here, k_1, k_2, and k_3 denote the radial distortion coefficients, and p_1 and p_2 represent the tangential distortion coefficients.

To reveal the co-registration accuracy between the aerial and terrestrial datasets, some checkpoints visible in the images captured by the two platforms were selected, measured, and triangulated. For precision, each checkpoint is measured in four images from the same platform at least. Finally, the 3D residuals between the triangulated position of the same checkpoint from the two platforms were computed, and the statistical parameters of mean value, maximum value, variance, and root mean square error (RMSE) were calculated.

3.2. Aerial–Terrestrial Datasets Used in the Experiments

To explore the performance of the different SCBA strategies, seven datasets were used in the experiments, including the benchmark datasets released by the International Society for Photogrammetry and Remote Sensing [47], SWJTU [21], and collected images. Tables 1 and 2 present some general information on the used image datasets.

Table 1. Basic information of the test datasets used in the study (GSD: ground sample distance).

Datasets	Sensors		Image Numbers		Average GSD (cm)	
	Aerial	Terrestrial	Aerial	Terrestrial	Aerial	Terrestrial
Rathaus	Sony Nex-7	Sony Nex-7	146	204	1.10	0.53
Stadthaus	Sony Nex-7	Canon 600D	345	132	1.79	0.52
Pferdestall	Sony Nex-7	Sony Nex-7	147	172	0.80	0.32
Verwaltung	Sony Nex-7	Canon 600D	728	351	0.65	0.28
Lohnhalle	Sony Nex-7	Canon 600D	426	194	0.66	0.27
SWJTU-LIB	Sony ILCE-5100	Canon M6	123	43	1.69	1.06
SZU-YPS	DJI FC6310R	DJI FC6310	178	53	2.47	0.86

Table 2. Basic information of the cameras used in the study.

Sensors	Resolution	Focal Length (mm/pixel)	Pixel Size (µm)
Sony Nex-7	6000 × 4000	16/4000	4 × 4
Canon 600D	5184 × 3456	20/4545	4.4 × 4.4
Sony ILCE-5100	6000 × 4000	40/10,256	3.9 × 3.9
Canon M6	6000 × 4000	18/4839	3.72 × 3.72
DJI FC6310R	5472 × 3648	8.8/3651	2.41 × 2.41

The Rathaus, Stadthaus, Pferdestall, Verwaltung, and Lohnhalle datasets were collected from Dortmund, Germany; the SWJTU-LIB dataset was collected from Chengdu, China; the SZU-YPS dataset was captured from Shenzhen, China. In all the tested datasets, most of the aerial (UAV) images are oblique, which would not only broaden the overlapping areas with the terrestrial views but also increase the intersection angles and result in better geometry accuracy for the image blocks. Sample images of the test datasets are shown in Figure 2. As shown in each row of Figure 2, although obvious perspective distortions and scale variations exist between aerial and terrestrial views, the lighting condition between them are rather similar. Therefore, the cross-platform image feature matching is more reliable than on those images with distinct radiometric discrepancies, which creates a good foundation for our cross-platform BA tests.

The UAV images in the SWJTU-LIB dataset were collected during regular strip flights, while the other aerial (UAV) and terrestrial images were captured surrounding and focusing on target buildings. In all datasets, the GPS information stored in the flights provided absolute ground control, and no ground control points are incorporated in the image orientation process. The SIFT algorithm [48] is implemented to find corresponding points between overlapping images, and the Gauss–Newton method [8] is adopted to solve the nonlinear least square problem in BA.

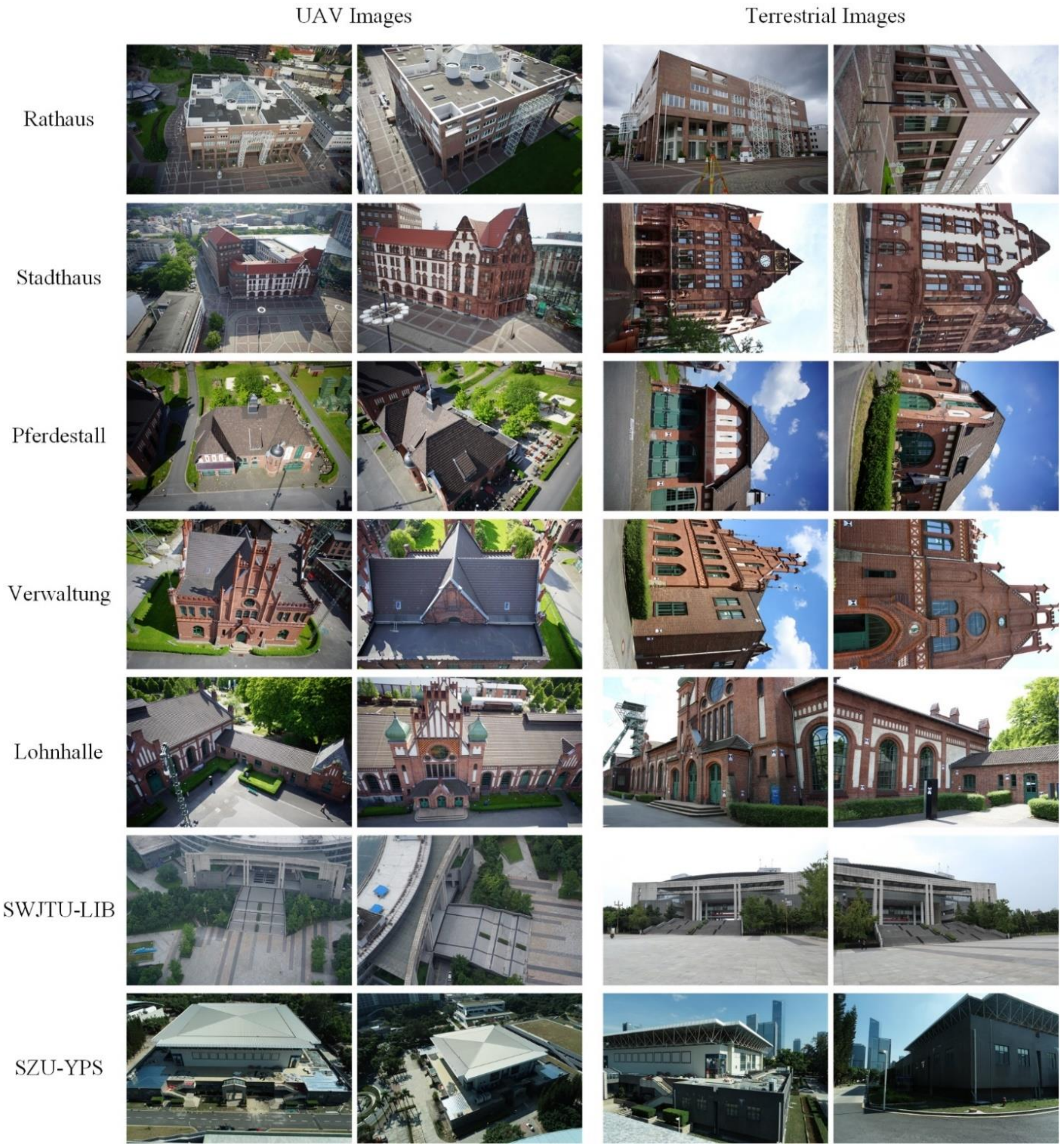

Figure 2. Sample images of the tested datasets.

4. Experimental Results and Analysis

4.1. Experimental Results

The EOPs and sparse point clouds after the BA of both aerial and terrestrial images are shown in Figure 3. In the seven tested areas, aerial and terrestrial images were correctly aligned, which enabled integrated 3D mapping and 3D scene reconstruction. Except for the Stadthaus dataset, the aerial images were captured through convergent shooting strategies, which resulted in stable image networks. Moreover, the terrestrial image networks in the Rathaus and Verwaltung datasets were more stable than those in the other five datasets.

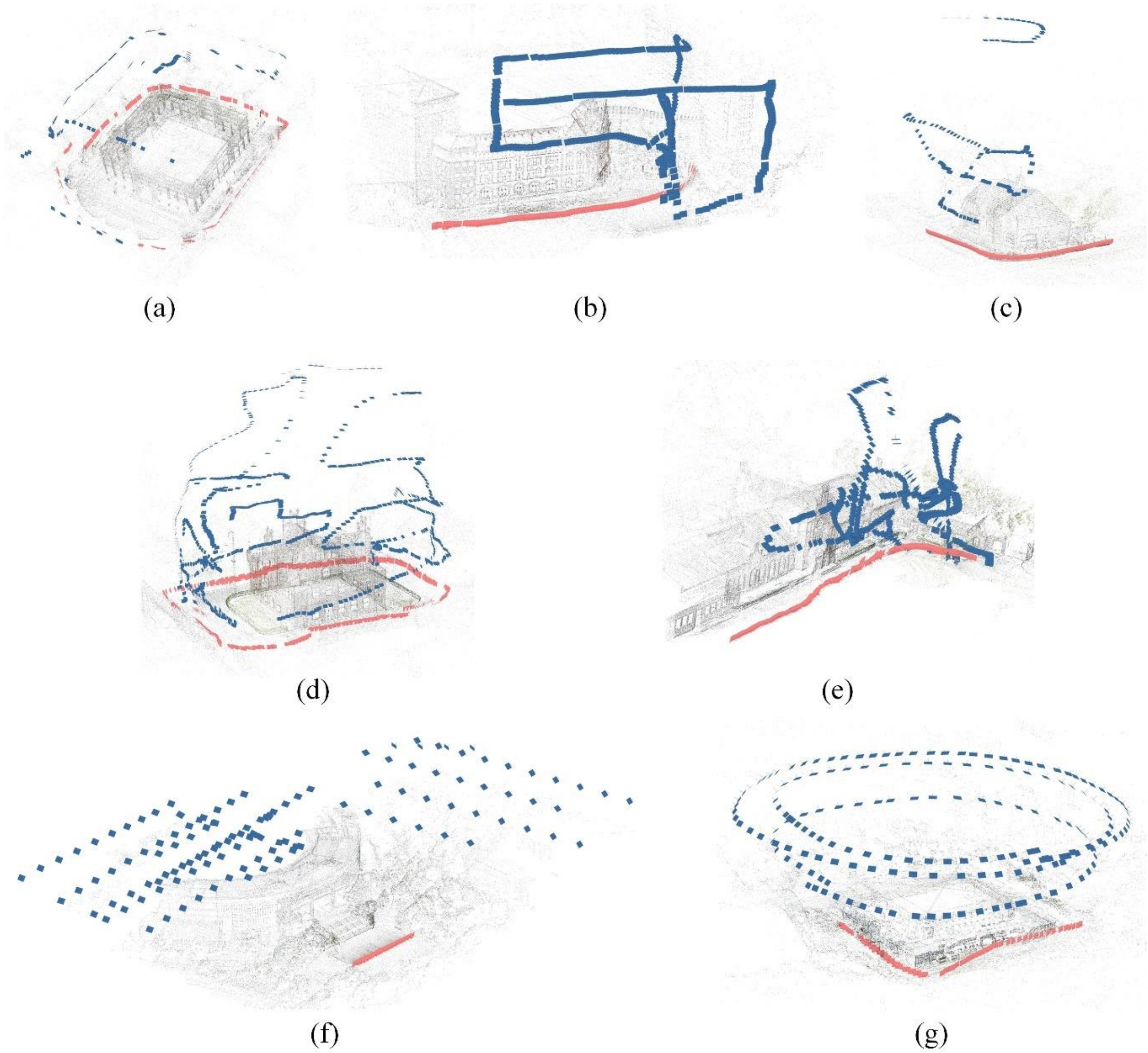

Figure 3. EOPs of images and sparse point clouds after aerial–terrestrial BA in the tested datasets. The blue and pink squares represent aerial and terrestrial images, respectively. (**a**) Rathaus; (**b**) Stadthaus; (**c**) Pferdestall; (**d**) Verwaltung; (**e**) Lohnhalle; (**f**) SWJTU-LIB; (**g**) SZU-YPS.

The refined IOPs of the aerial (UAV) and terrestrial images after SCBA, and the IOPs after aerial–terrestrial BA using the four different strategies, are presented in Tables 3–9. For convenience, the focal lengths and coordinates of the principal points are represented in pixels. As presented in Tables 3–9, during the refining of the IOPs of aerial images in the BA, the focal lengths after optimization were highly stable, and the differences in focal length between RA_RT, RA_FT, and FA_FT&FA_RT were within 1 pixel. Meanwhile, the corresponding differences for terrestrial images were slightly greater; for example, the focal length after FA_RT BA strategies was greater than the original value of 1 pixel (Table 5). The principal points x_p and y_p also exhibited the same trend. The differences in x_p and y_p between the four compared BA strategies were less than 2 pixels in all of the tested datasets for aerial images, but greater than 2 pixels in the tested datasets for terrestrial images. For the Pferdestall dataset, the variations between x_p obtained through the FA_RT and FA_FT&RA_FT methods were greater than 3 pixels. For the other additional parameters related to radial distortion and tangential distortion, the adjusted values obtained through the four compared BA strategies were similar.

Table 3. IOPs of aerial (UAV) and terrestrial images after the BA of the Rathaus dataset using different SC strategies.

IOPs	Aerial (UAV)			Terrestrial		
	FA_FT & FA_RT	RA_RT	RA_FT	FA_FT & RA_FT	RA_RT	FA_RT
f (pixels)	4068.87807	4068.08033	4068.00866	4067.36138	4067.21119	4067.21142
x_p (pixels)	−19.5567	−19.3128	−19.3223	−19.3638	−20.6114	−20.7009
y_p (pixels)	33.0412	33.2878	33.3547	11.1559	10.0832	10.0046
k_1	−0.065801	−0.0658264	−0.0658341	−0.0658597	−0.0659561	−0.0659577
k_2	0.0875099	0.0874178	0.0874218	0.090855	0.0910229	0.0910422
k_3	0.0104374	0.0104204	0.0104062	0.00545872	0.00526018	0.0052374
p_1 (10^{-3})	−0.374962	−0.364741	−0.365756	−0.581976	−0.65487	−0.661701

Table 4. IOPs of aerial (UAV) images and terrestrial images after the BA of the Stadthaus dataset using different SC strategies.

IOPs	Aerial (UAV)			Terrestrial		
	FA_FT & FA_RT	RA_RT	RA_FT	FA_FT & RA_FT	RA_RT	FA_RT
f (pixels)	4068.23085	4068.30545	4068.30622	4771.20366	4771.37552	4771.37324
x_p (pixels)	−18.2148	−18.2052	−18.2064	25.5432	25.0831	25.0884
y_p (pixels)	32.6839	32.6683	32.6662	5.61209	6.22631	6.2269
k_1	−0.0674332	−0.0673206	−0.0673213	−0.0980488	−0.0981193	−0.0981196
k_2	0.0917847	0.0914299	0.0914315	0.101131	0.101545	0.101546
k_3	0.00715928	0.00748223	0.00748108	−0.0294874	−0.0301175	−0.0301188
p_1 (10^{-3})	−0.285357	−0.276588	−0.276665	0.161089	0.13874	0.138956

Table 5. IOPs of aerial (UAV) and terrestrial images after the BA of the Pferdestall dataset using different SC strategies.

IOPs	Aerial (UAV)			Terrestrial		
	FA_FT & FA_RT	RA_RT	RA_FT	FA_FT & RA_FT	RA_RT	FA_RT
f (pixels)	4067.14882	4066.88138	4066.84198	4063.41961	4064.35911	4064.43471
x_p (pixels)	−17.8754	−18.788	−18.8207	−27.1081	−30.1829	−30.3601
y_p (pixels)	33.7612	33.809	33.8791	23.0151	22.4765	22.3341
k_1	−0.0632268	−0.0634589	−0.0634678	−0.0631941	−0.0631213	−0.0630996
k_2	0.0852705	0.0857281	0.0857319	0.082288	0.0828569	0.0828302
k_3	0.0116867	0.0111896	0.0111769	0.0117249	0.0110463	0.0110572
p_1 (10^{-3})	−0.294877	−0.334246	−0.336489	−0.712583	−0.871517	−0.88116

Table 6. IOPs of aerial (UAV) and terrestrial images after the BA of the Verwaltung dataset using different SC strategies.

IOPs	Aerial (UAV)			Terrestrial		
	FA_FT & FA_RT	RA_RT	RA_FT	FA_FT & RA_FT	RA_RT	FA_RT
f (pixels)	4066.96303	4066.85896	4066.78135	4771.08311	4771.73777	4771.75601
x_p (pixels)	−17.6385	−18.7535	−18.9217	39.2519	40.2749	40.3715
y_p (pixels)	32.9631	32.9664	32.9728	15.5072	18.9996	19.3648
k_1	−0.0630105	−0.0630302	−0.0630428	−0.0972711	−0.097242	−0.0972324
k_2	0.0855557	0.0855932	0.085621	0.0991191	0.0994272	0.0994101
k_3	0.0110263	0.0109472	0.0109034	−0.0268749	−0.0275584	−0.027555
p_1 (10^{-3})	−0.318407	−0.397235	−0.408668	0.209112	0.247171	0.249551

Table 7. IOPs of aerial (UAV) and terrestrial images after the BA of the Lohnhalle dataset using different SC strategies.

IOPs	Aerial (UAV)			Terrestrial		
	FA_FT & FA_RT	**RA_RT**	**RA_FT**	**FA_FT & RA_FT**	**RA_RT**	**FA_RT**
f (pixels)	4066.36477	4066.67647	4066.63846	4771.68997	4771.64815	4771.59545
x_p (pixels)	−15.4462	−17.2505	−17.2227	36.5568	36.7277	37.4373
y_p (pixels)	32.7025	33.2739	33.3615	9.8315	9.11882	8.9211
k_1	−0.0629404	−0.0632259	−0.0632351	−0.097497	−0.0972915	−0.0973073
k_2	0.0841735	0.0848282	0.084834	0.0999886	0.0987108	0.0987518
k_3	0.0124227	0.0119863	0.0119737	−0.0284858	−0.0259536	−0.0260674
p_1 (10^{-3})	−0.195024	−0.273688	−0.273387	0.16963	0.208466	0.229709

Table 8. IOPs of aerial (UAV) and terrestrial images after the BA of the SWJTU-LIB dataset using different SC strategies.

IOPs	Aerial (UAV)			Terrestrial		
	FA_FT & FA_RT	**RA_RT**	**RA_FT**	**FA_FT & RA_FT**	**RA_RT**	**FA_RT**
f (pixels)	10,237.6675	10,237.9058	10,237.9642	4941.57027	4942.06517	4942.13066
x_p (pixels)	42.6596	42.6941	42.6453	−0.888504	−1.40598	−1.26491
y_p (pixels)	−40.7954	−40.9146	−40.946	−6.8541	−7.5046	−7.93967
k_1	0.0635718	0.0635644	0.0635714	−0.0965242	−0.0966374	−0.0966099
k_2	−0.0209569	−0.0199052	−0.0198847	0.12534	0.125812	0.125739
k_3	−0.251215	−0.258173	−0.258564	−0.0311879	−0.0314217	−0.0313546
p_1 (10^{-3})	0.0387311	0.0397392	0.0377901	0.662126	0.626801	0.63709

Table 9. IOPs of aerial (UAV) and terrestrial images after the BA of the SZU-YPS dataset using different SC strategies.

IOPs	Aerial (UAV)			Terrestrial		
	FA_FT & FA_RT	**RA_RT**	**RA_FT**	**FA_FT & RA_FT**	**RA_RT**	**FA_RT**
f (pixels)	3695.45203	3695.44054	3695.44076	3694.52433	3694.80362	3694.80258
x_p (pixels)	−6.99241	−6.95527	−6.95292	−3.48738	−4.06529	−4.06683
y_p (pixels)	−13.3504	−13.3379	−13.3402	−14.6709	−14.0835	−14.0841
k_1	−0.0146416	−0.014608	−0.0146078	−0.0130628	−0.013033	−0.0130332
k_2	0.00491831	0.00485732	0.00485649	9.75272×10^{-5}	8.4753×10^{-5}	8.48471×10^{-5}
k_3	0.00461913	0.00465268	0.00465336	0.00915655	0.00924262	0.00924254
p_1 (10^{-3})	−0.852657	−0.850009	−0.849824	−0.545586	−0.583879	−0.584001

The adjusted values of RA_FT and RA_RT for the aerial images were closer to each other than they were to the original values that were only adjusted in the previous SCBA. Consequently, the optimized IOPs of FA_RT and RA_RT for the terrestrial images were closer to each other than they were to the optimized IOPs of FA_FT&RA_FT. As given in Table 1, the UAV images and terrestrial images in Rathaus and Pferdestall were obtained using the same cameras. However, the obtained IOPs were closer to one another in the Rathaus dataset than in the Pferdestall dataset (Tables 3 and 5).

To verify the performance of co-registration between aerial and terrestrial images processed using the four compared BA strategies, the triangulated 3D positions of some checkpoints that were visible in both the aerial and terrestrial images were compared. Figure 4 illustrates some samples of checkpoints measured in the tested datasets. These points are measured on targets which are intentionally put by the data collectors, and the other CPs are measured on distinct corners in the scenes. For each checkpoint, at least four observations were measured in images captured by one platform, and the maximum back projection errors for checkpoints are limited to one pixel. Eleven to seventeen checkpoints were extracted in each of the seven tested datasets. Most of the checkpoints are distributed on vertical walls because those are the major common visible areas for aerial and terrestrial views. To reach even distribution as fair as possible in both planar and vertical directions, there are also some checkpoints measured on the roof corners and the ground. However, the absolute 3D position for checkpoints has not been measured by any topographic support since this study focuses on the relative cross-platform co-registration accuracy.

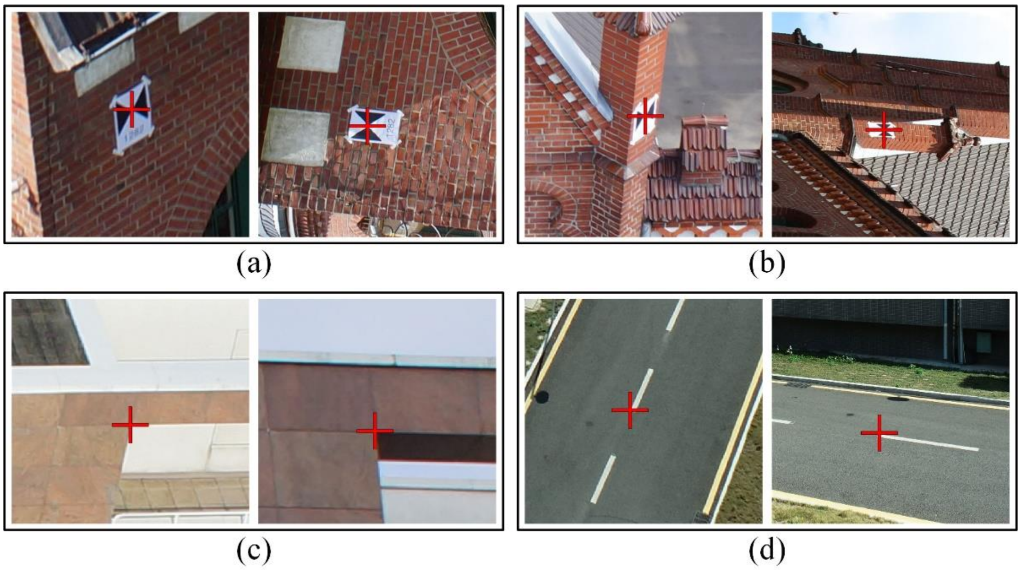

Figure 4. Samples of checkpoints measured in the tested datasets. In each subgraph, the left is measured in aerial (UAV) image while the right is measured in terrestrial image. (**a**,**b**), checkpoints measured on targets; (**c**,**d**), checkpoints measured on distinct corners.

The statistics of 3D errors at checkpoints are listed in Table 10. For the seven tested areas, the RMSEs in the X, Y, and Z directions ranged from 4.448 to 202.729 mm. The minimax RMSE values occurred in the Verwaltung dataset, and the maximum values occurred in the SWJTU_LIB dataset. This was possibly due to the occurrence of different image GSDs and image numbers (Table 1). Comparison of the results of different BA strategies revealed that the FA_FT and FA_RT methods exhibited the minimum RMSE values. In Rathaus, Stadthaus, Pferdestall, SWJTU-LIB, and SZU-YPS, the FA_FT method obtained the minimum 3D RMSE values and the minimax deviation values, which corresponded to the highest co-registration accuracies between aerial and terrestrial images.

To qualitatively reveal the relative co-registration accuracies of the different BA strategies, the relative RMSE (rRMSE) was calculated using Equation (4).

$$\mathrm{rRMSE_i} = \frac{\mathrm{RMSE_i}}{\frac{1}{n}\sum_{j=1}^{n} \mathrm{RMSE_j}} * 100 \tag{4}$$

In Equation (4), $\mathrm{RMSE_i}$ denotes the calculated absolute RMSE value of the current method; n denotes the total number of compared methods. The lower the rRMSE value, the higher the relative co-registration accuracy. Moreover, the mean rRMSE value remained 100 in one test area. The calculated rRMSEs of the seven tested areas calculated using this equation are shown in Figure 5. For the seven tested areas, the rRMSEs obtained using the FA_FT method were less than the mean values. For the Stadthaus and SZU-YPS datasets, the rRMSEs obtained using the FA_FT method were remarkably greater than those obtained using the other three methods. For all datasets except SWJTU-LIB, the rRMSEs obtained using the FA_RT and RA_RT methods were similar. The lowest rRMSEs obtained through FA_RT belonged to the Verwaltung and Lohnhalle datasets, and the RA_RT method obtained unstable results for the SWJTU-LIB dataset. The RA_FT method obtained the largest rRMSEs for four test areas, namely Rathaus, Stadthaus, Verwaltung, and Lohnhalle. For the Lohnhalle dataset, the rRMSE values of the four methods were the closest.

Table 10. Statistics of checkpoints. No. CPs: number of checkpoints. RMSE: root mean square error. The values in bold indicate the minimal values in the compared four methods.

Dataset	No. CPs	Method	X RMSE (mm)	Y RMSE (mm)	Z RMSE (mm)	RMSE (mm)	Maximum (mm)
Rathaus	15	FA_FT	43.397	10.201	13.236	**46.503**	**123.952**
		FA_RT	43.574	10.990	13.035	46.791	125.882
		RA_RT	43.522	11.082	12.950	46.741	125.408
		RA_FT	43.682	10.750	13.377	46.932	124.326
Stadthaus	15	FA_FT	32.977	7.701	21.708	**40.224**	**78.887**
		FA_RT	32.866	7.337	24.767	41.802	83.446
		RA_RT	32.828	7.310	24.773	41.771	83.240
		RA_FT	32.545	7.988	25.741	42.256	85.076
Pferdestall	13	FA_FT	17.963	23.083	22.517	**36.912**	**85.580**
		FA_RT	15.362	25.831	22.761	37.700	91.189
		RA_RT	15.110	25.989	22.787	37.722	91.718
		RA_FT	17.988	23.220	22.474	36.984	85.799
Verwaltung	19	FA_FT	4.455	2.315	4.586	6.800	18.916
		FA_RT	4.448	2.434	4.445	**6.743**	**18.488**
		RA_RT	4.536	2.388	4.450	6.788	19.376
		RA_FT	4.572	2.270	4.598	6.870	20.166
Lohnhalle	17	FA_FT	9.090	7.920	36.324	38.272	148.034
		FA_RT	9.157	7.877	36.244	**38.203**	**147.707**
		RA_RT	9.127	7.887	36.346	38.296	148.038
		RA_FT	9.173	7.972	36.376	38.353	148.229
SWJTU-LIB	11	FA_FT	120.657	200.995	57.237	**241.316**	**417.162**
		FA_RT	120.999	201.168	57.379	241.664	419.108
		RA_RT	116.547	202.729	118.976	262.369	414.491
		RA_FT	120.019	201.469	57.030	241.343	417.530
SZU-YPS	14	FA_FT	10.776	10.059	14.005	**20.333**	**39.278**
		FA_RT	13.299	9.548	14.179	21.659	46.281
		RA_RT	13.293	9.558	14.141	21.634	46.215
		RA_FT	12.601	9.998	14.388	21.582	47.669

4.2. Experimental Analysis

To further investigate the BA results of aerial and terrestrial images for different SC strategies, the tie point distribution in the image space was plotted and analyzed. Because an even tie point distribution is favorable to the SC process, the uniformities of both aerial and terrestrial tie points were qualitatively and quantitatively evaluated. The image space was equally divided into 100 parts in both the horizontal and vertical directions, and 10,000 grids were created. The number of tie points that belonged to each grid was obtained from the image coordinates of tie points.

These numbers were plotted using heatmaps to visually reveal the tie point distribution for aerial and terrestrial images. The mean values, standard deviations, and normalized standard deviations of the numbers are presented in Table 11. To compare the distribution between different image platforms and datasets, the normalized number of tie points (norNTP) was calculated using Equation (5).

$$\mathrm{norNTP_i} = \frac{\mathrm{NTP_i}}{\frac{1}{n}\sum_{j=1}^{n}\mathrm{NTP_j}} * 10,000 \quad (5)$$

where NTP_i denotes the number of tie points in a certain grid, and $norNTP_i$ is the normalized value. Thus, the mean value of the normalized number of tie points in each grid was 1. The tie point distributions are illustrated in Figures 6–12 using heatmaps generated from the norNTP values and histograms.

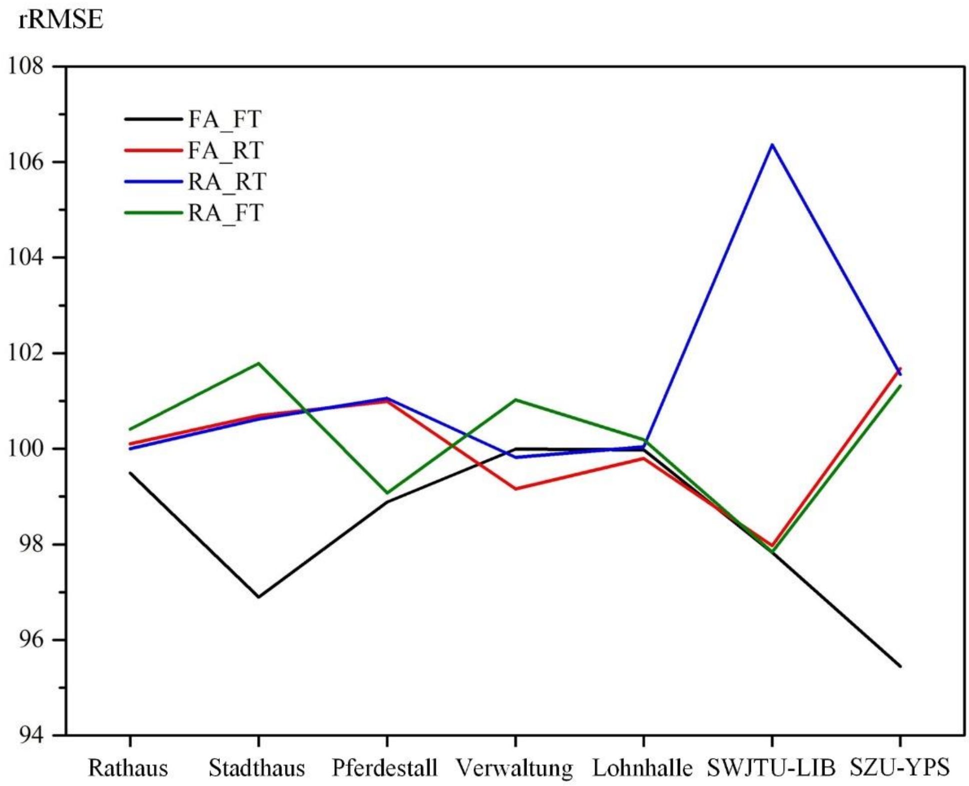

Figure 5. The rRMSEs of the four compared BA strategies used for the seven tested areas. Different colors indicate different BA methods; black: FA_FT; red: FA_RT; blue: RA_RT; green: RA_FT.

Table 11. Mean values (MVs), standard deviations (STDs), and normalized standard deviations (NSTDs) of the number of points in equant grids.

Datasets	MVs		STDs		NSTDs	
	Aerial	**Terrestrial**	**Aerial**	**Terrestrial**	**Aerial**	**Terrestrial**
Rathaus	43.655	51.902	11.280	21.967	0.258	0.423
Stadthaus	60.061	68.535	16.972	25.363	0.283	0.370
Pferdestall	36.058	40.351	9.566	27.497	0.265	0.681
Verwaltung	184.195	100.603	38.425	43.800	0.209	0.435
Lohnhalle	72.219	40.932	15.342	14.088	0.212	0.344
SWJTU-LIB	41.325	11.851	10.286	8.665	0.249	0.731
SZU-YPS	38.710	11.930	13.881	7.671	0.359	0.643

The mean values, standard deviations, and NSTDs of the number of points in equant grids for the tested datasets are given in Table 11. Comparison of the NSTDs between different platforms revealed that in the seven tested datasets, the NSTDs for the aerial images were considerably smaller than those for terrestrial images, indicating that the tie point distribution in the aerial platforms was more uniform than that in the terrestrial platforms.

The NSTD values are consistent with the visual expression of the results in Figures 6–12. As illustrated in Figures 6a, 7a, 8a, 9a, 10a, 11a and 12a, the tie points in the aerial images were mostly distributed around the expected mean value (1.0). However, the corresponding terrestrial images (Figures 6b, 7b, 8b, 9b, 10b, 11b and 12b) featured blue areas, which indicate fewer tie points or even the absence of tie point spread in the margin of one side or a corner of the images. Moreover, owing to the differences in scene content, the images for the Pferdestall, SWJIT-LIB, and SZU-YPS datasets also featured large red areas (Figures 8b, 11b and 12b), which indicates that the tie points of the terrestrial images of the datasets had distinct focus points.

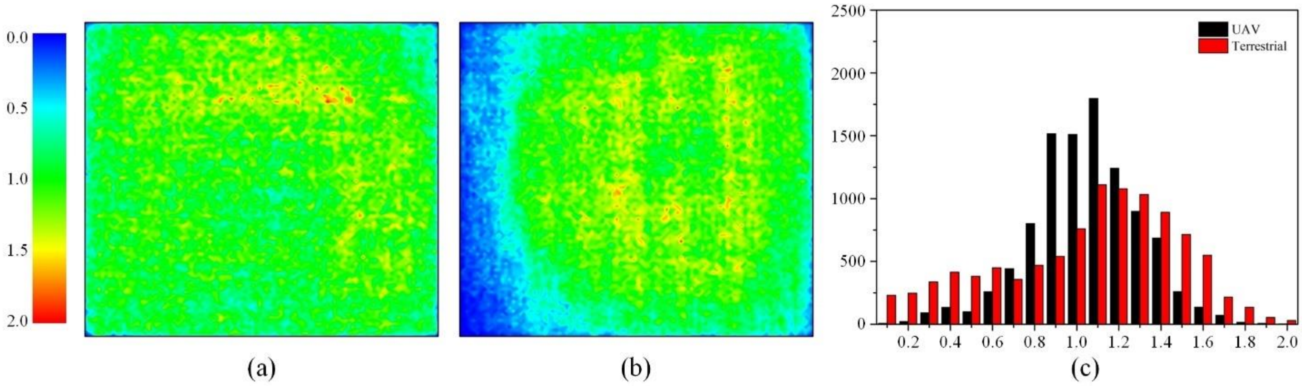

Figure 6. Tie point distributions for the Rathaus dataset: (**a**) aerial (UAV) and (**b**) terrestrial image spaces; (**c**) histograms.

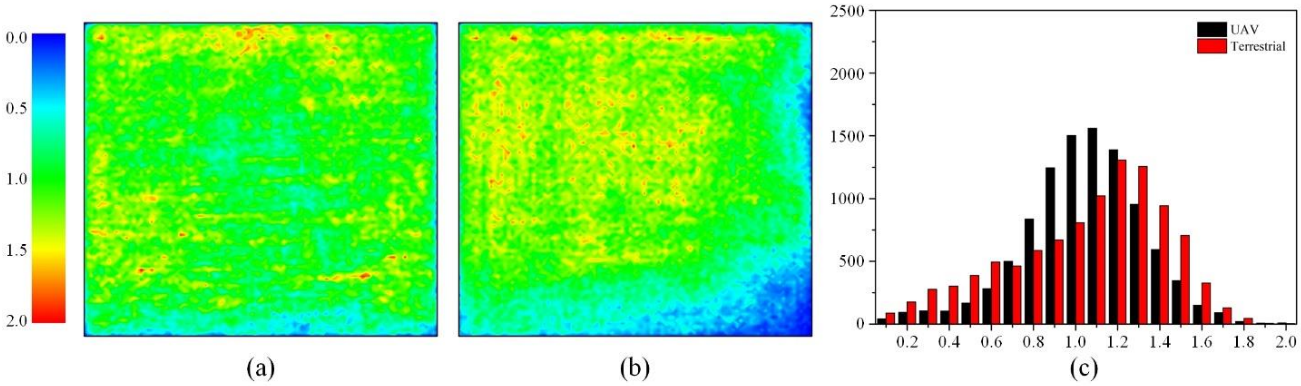

Figure 7. Tie point distributions for the Stadthaus dataset: (**a**) aerial (UAV) and (**b**) terrestrial image spaces; (**c**) histograms.

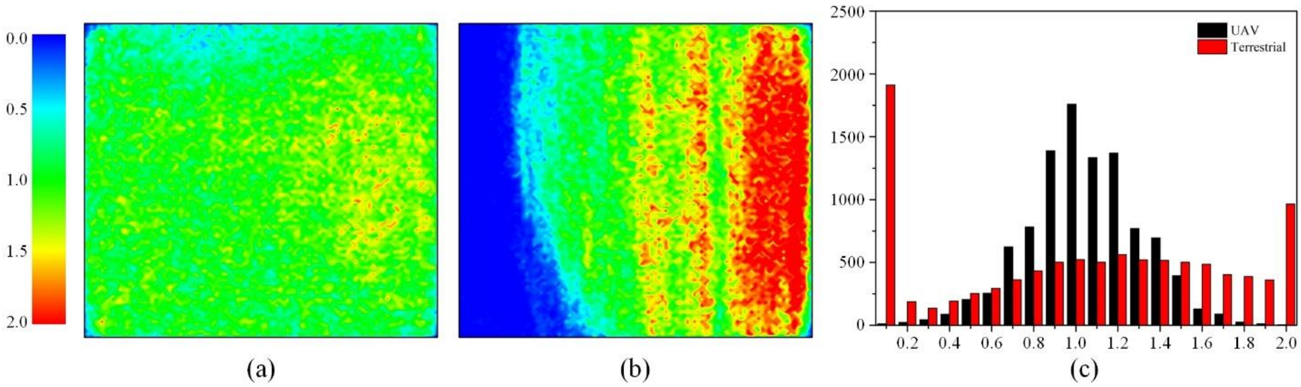

Figure 8. Tie point distributions for the Pferdestall dataset: (**a**) aerial (UAV) and (**b**) terrestrial image spaces; (**c**) histograms.

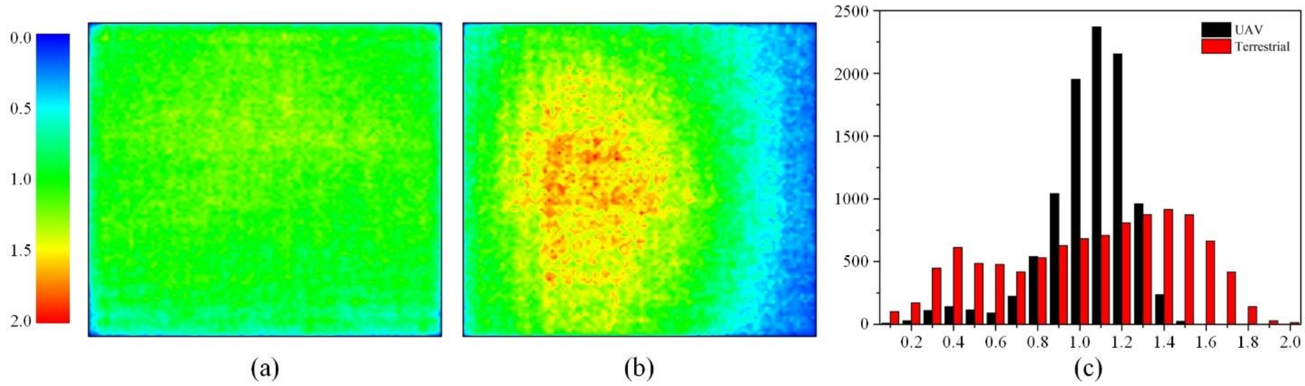

Figure 9. Tie point distributions for the Verwaltung dataset; (**a**) aerial (UAV) and (**b**) terrestrial image spaces; (**c**) histograms.

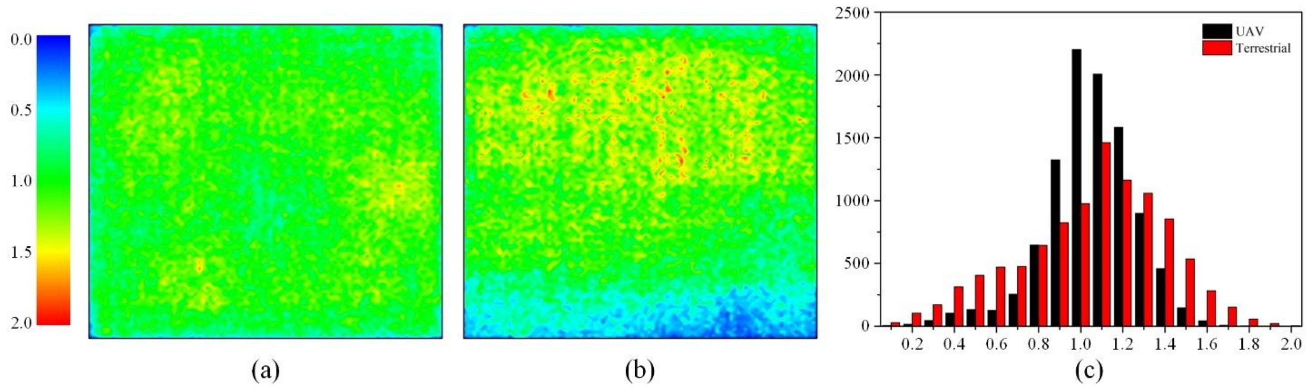

Figure 10. Tie point distributions for the Lohnhalle dataset: (**a**) aerial (UAV) and (**b**) terrestrial image spaces; (**c**) histograms.

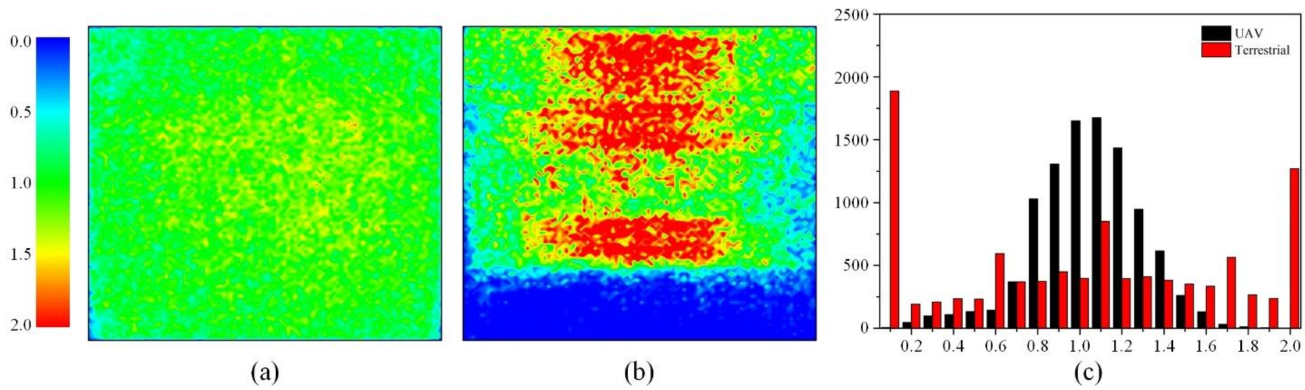

Figure 11. Tie point distributions for the SWJIT-LIB dataset: (**a**) aerial (UAV) and (**b**) terrestrial image spaces; (**c**) histograms.

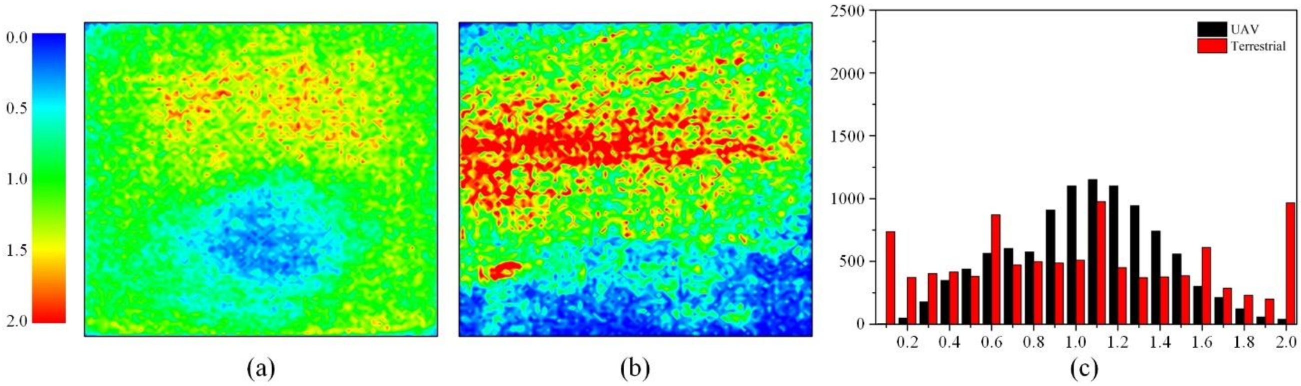

Figure 12. Tie point distributions for the SZU-YPS dataset: (**a**) aerial (UAV) and (**b**) terrestrial image spaces; (**c**) histograms.

In an image space with perfectly distributed tie points, the histograms of NSTDs will have large values of bands approximately 1.0, and low values (or even zero) at other bands, particularly the minimum and maximum bands. In the histograms (Figures 6c, 7c, 8c, 9c, 10c, 11c and 12c), the peaks for aerial images were all at 1.0 and 1.1 for the seven tested areas, while the peaks for terrestrial images varied. For the Verwaltung and the Lohnhalle datasets, the histogram peaks for aerial images were as large as 2000, and the norNTPs for more than 7000 grids were between 0.9 and 1.2, which represents a favorable tie point distribution for SCBA. Meanwhile, the histograms for terrestrial images were rather mild. The histogram peaks for terrestrial images of the Rathaus, Stadthaus, Verwaltung, and Lohnhalle datasets were between 1.1 and 1.4. For the other three datasets, the histograms featured high values of columns for the minimum and the maximum bands for terrestrial images, which suggests extremely uneven tie point distribution. In the Pferdestall and SWJIT-LIB datasets, the norNTPs of approximately 1800 grid points were less than 0.1, implying that no tie points fell in approximately 18% of the image space. Furthermore, except for the Stadthaus dataset, the norNTPs for aerial images between 0.0 to 0.1 were near-zero, suggesting that the tie points almost covered the whole images. Moreover, for the Stadthaus, Verwaltung, and Lohnhalle datasets, the norNTPs for terrestrial images between 0.0 and 0.1 were also small, which means that the tie points were rather comprehensively distributed in the image space.

5. Discussion

According to the experimental results and statistical analysis, Table 12 summarizes the network stability, tie point distribution uniformity, and best SC strategies for the seven tested datasets.

Table 12. Summary of the network stability (NS), distribution uniformity of tie points (DUTPs), and best self-calibration strategy (BSCS) in the BA of aerial–terrestrial images for the seven datasets.

Datasets	NS		DUTPs		BSCS	
	Aerial	**Terrestrial**	**Aerial**	**Terrestrial**	**Aerial**	**Terrestrial**
Rathaus	Yes	Yes	Good	Fair	Fix	Fix
Stadthaus	No	No	Fair	Fair	Fix	Fix
Pferdestall	Yes	No	Good	Bad	Fix	Fix
Verwaltung	Yes	Yes	Good	Fair	Fix	Refine
Lohnhalle	Yes	No	Good	Fair	Fix	Refine
SWJTU-LIB	Yes	No	Good	Bad	Fix	Fix
SZU-YPS	Yes	No	Fair	Bad	Fix	Fix

Because the tie point distribution in aerial images is fairly even, the BA of aerial and terrestrial images may not require a second round of SC for aerial cameras. Moreover, fixing the IOPs of aerial cameras (which have already been refined) during the BA could reduce the number of unknown parameters.

However, this is not the case for terrestrial cameras. Owing to the different shooting conditions, the tie point coverage in terrestrial images is insufficient to regain the physical distortion parameters through theoretical lens calibration. Thus, herein, the IOPs obtained in the first round of the SCBA of terrestrial images inadequately represented the real IOPs of the terrestrial cameras. Hence, the IOPs of terrestrial images can be refined in the second-round BA that combines both aerial and terrestrial images if the tie points have relatively even distribution and large format coverage.

Considering this assumption, the FA_RT method will provide the best co-registration results for the Stadthaus, Verwaltung, and Lohnhalle datasets. However, the minimum rRMSE values acquired through the FA_FT strategy belonged to the Stadthaus dataset, presumably because the networks of aerial images in the Stadthaus dataset were not stable. Therefore, refining the IOPs of terrestrial cameras may degrade the EOPs of aerial images and result in suboptimal cross-platform co-registration accuracy.

According to the experimental results and above analysis, some suggestions regarding the SC strategies in the BA of aerial and terrestrial image blocks are offered. First, for aerial images, with better tie point distribution than terrestrial images, it is better to fix the IOPs in the cross-platform BA. Second, for most cases, fixing the IOPs of terrestrial images in the second-round BA will improve the co-registration accuracy. Third, if the tie point distribution in terrestrial images is relatively even and comprehensive and the networks of aerial images are reasonably stable, refining the IOPs of terrestrial cameras in the BA may yield the best results.

Future tests will investigate the effect of SC strategies on mathematical lens distortion parameters such as the Fourier SC additional parameters [34] in cross-platform image BA. Moreover, automatic optimal SC strategy selection methods related to cross-platform image orientation should be developed and investigated, to build unified precision 3D mapping references for multi-platform photogrammetry.

Author Contributions: Conceptualization, L.X. and W.W.; methodology, L.X. and Q.Z.; software, L.X. and H.H.; validation, X.Y., Y.Z. and X.Q.; formal analysis, Q.Z. and X.L.; supervision, R.G. All authors wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by National Key R&D Program of China (2019YFB2103104), the National Natural Science Foundation of China (42001407, 42104012, 41971341, 41971354), the Guangdong Basic and Applied Basic Research Foundation (2019A1515110729, 2019A1515111163, 2019A1515010748, 2019A1515011872), the Open Research Fund of State Key Laboratory of Information Engineering in Surveying Mapping and Remote Sensing, Wuhan University (20E02, 20R06), the Guangdong Science and Technology Strategic Innovation Fund (the Guangdong–Hong Kong–Macau Joint Laboratory Program, 2020B1212030009), and the Shenzhen Key Laboratory of Digital Twin Technologies for Cities (ZDSYS20210623101800001).

Data Availability Statement: The datasets used in this study are partially available in [21,47].

Acknowledgments: The authors gratefully acknowledge the provision of the datasets by ISPRS and EuroSDR, released in conjunction with the ISPRS Scientific Initiatives 2014 and 2015, led by ISPRS ICWG I/Vb.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Jiang, S.; Jiang, W.; Wang, L. Unmanned Aerial Vehicle-Based Photogrammetric 3D Mapping: A Survey of Techniques, Applications, and Challenges. *IEEE Geosci. Remote Sens. Mag.* **2022**, 2–38. [CrossRef]
2. Bouzas, V.; Ledoux, H.; Nan, L. Structure-aware Building Mesh Polygonization. *ISPRS J. Photogramm. Remote Sens.* **2020**, *167*, 432–442. [CrossRef]
3. Jiang, S.; Jiang, C.; Jiang, W. Efficient structure from motion for large-scale UAV images: A review and a comparison of SfM tools. *ISPRS J. Photogramm. Remote Sens.* **2020**, *167*, 230–251. [CrossRef]

4. Rupnik, E.; Nex, F.; Toschi, I.; Remondino, F. Aerial multi-camera systems: Accuracy and block triangulation issues. *ISPRS J. Photogramm. Remote Sens.* **2015**, *101*, 233–246. [CrossRef]
5. Rau, J.; Jhan, J.; Hsu, Y. Analysis of Oblique Aerial Images for Land Cover and Point Cloud Classification in an Urban Environment. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 1304–1319. [CrossRef]
6. Toschi, I.; Ramos, M.M.; Nocerino, E.; Menna, F.; Remondino, F.; Moe, K.; Poli, D.; Legat, K.; Fassi, F. Oblique photogrammetry supporting 3d urban reconstruction of complex scenarios. *ISPRS-Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2017**, *XLII-1/W1*, 519–526. [CrossRef]
7. Zhou, G.; Bao, X.; Ye, S.; Wang, H.; Yan, H. Selection of Optimal Building Facade Texture Images From UAV-Based Multiple Oblique Image Flows. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 1534–1552. [CrossRef]
8. Triggs, B.; McLauchlan, P.F.; Hartley, R.I.; Fitzgibbon, A.W. *Bundle Adjustment—A Modern Synthesis*; Springer: Berlin/Heidelberg, Germay, 2000; pp. 153–177.
9. Honkavaara, E.; Ahokas, E.; Hyyppä, J.; Jaakkola, J.; Kaartinen, H.; Kuittinen, R.; Markelin, L.; Nurminen, K. Geometric test field calibration of digital photogrammetric sensors. *ISPRS J. Photogramm. Remote Sens.* **2006**, *60*, 387–399. [CrossRef]
10. Hu, H.; Zhu, Q.; Du, Z.; Zhang, Y.; Ding, Y. Reliable spatial relationship constrained feature point matching of oblique aerial images. *Photogramm. Eng. Remote Sensing.* **2015**, *81*, 49–58. [CrossRef]
11. Xie, L.; Hu, H.; Wang, J.; Zhu, Q.; Chen, M. An asymmetric re-weighting method for the precision combined bundle adjustment of aerial oblique images. *ISPRS J. Photogramm. Remote Sens.* **2016**, *117*, 92–107. [CrossRef]
12. Wang, X.; Xiao, T.; Kasten, Y. A hybrid global structure from motion method for synchronously estimating global rotations and global translations. *ISPRS J. Photogramm. Remote Sens.* **2021**, *174*, 35–55. [CrossRef]
13. Duane, C.B. Close-range camera calibration. *Photogramm. Eng.* **1971**, *37*, 855–866.
14. Fraser, C.S. Digital camera self-calibration. *ISPRS J. Photogramm. Remote Sens.* **1997**, *52*, 149–159. [CrossRef]
15. Lichti, D.D.; Kim, C.; Jamtsho, S. An integrated bundle adjustment approach to range camera geometric self-calibration. *ISPRS J. Photogramm. Remote Sens.* **2010**, *65*, 360–368. [CrossRef]
16. Luhmann, T.; Fraser, C.; Maas, H. Sensor modelling and camera calibration for close-range photogrammetry. *ISPRS J. Photogramm. Remote Sens.* **2016**, *115*, 37–46. [CrossRef]
17. Griffiths, D.; Burningham, H. Comparison of pre-and self-calibrated camera calibration models for UAS-derived nadir imagery for a SfM application. *Prog. Phys. Geogr. Earth Environ.* **2019**, *43*, 215–235. [CrossRef]
18. Babapour, H.; Mokhtarzade, M.; Valadan Zoej, M.J. Self-calibration of digital aerial camera using combined orthogonal models. *ISPRS J. Photogramm. Remote Sens.* **2016**, *117*, 29–39. [CrossRef]
19. Gao, X.; Shen, S.; Zhou, Y.; Cui, H.; Zhu, L.; Hu, Z. Ancient Chinese architecture 3D preservation by merging ground and aerial point clouds. *ISPRS J. Photogramm. Remote Sens.* **2018**, *143*, 72–84. [CrossRef]
20. Wu, B.; Xie, L.; Hu, H.; Zhu, Q.; Yau, E. Integration of aerial oblique imagery and terrestrial imagery for optimized 3D modeling in urban areas. *ISPRS J. Photogramm. Remote Sens.* **2018**, *139*, 119–132. [CrossRef]
21. Zhu, Q.; Wang, Z.; Hu, H.; Xie, L.; Ge, X.; Zhang, Y. Leveraging photogrammetric mesh models for aerial-ground feature point matching toward integrated 3D reconstruction. *ISPRS J. Photogramm. Remote Sens.* **2020**, *166*, 26–40. [CrossRef]
22. Migliazza, M.; Carriero, M.T.; Lingua, A.; Pontoglio, E.; Scavia, C. Rock Mass Characterization by UAV and Close-Range Photogrammetry: A Multiscale Approach Applied along the Vallone dell'Elva Road (Italy). *Geosciences* **2021**, *11*, 436. [CrossRef]
23. Wang, X.; Rottensteiner, F.; Heipke, C. Structure from motion for ordered and unordered image sets based on random k-d forests and global pose estimation. *ISPRS J. Photogramm. Remote Sens.* **2019**, *147*, 19–41. [CrossRef]
24. Clarke, T.A.; Fryer, J.G. The development of camera calibration methods and models. *Photogramm. Rec.* **1998**, *16*, 51–66. [CrossRef]
25. Snavely, N.; Seitz, S.M.; Szeliski, R. Modeling the World from Internet Photo Collections. *Int. J. Comput. Vis.* **2008**, *80*, 189–210. [CrossRef]
26. Martínez-Carricondo, P.; Agüera-Vega, F.; Carvajal-Ramírez, F.; Mesas-Carrascosa, F.; García-Ferrer, A.; Pérez-Porras, F. Assessment of UAV-photogrammetric mapping accuracy based on variation of ground control points. *Int. J. Appl. Earth Obs.* **2018**, *72*, 1–10. [CrossRef]
27. James, M.R.; Robson, S. Mitigating systematic error in topographic models derived from UAV and ground-based image networks. *Earth Surf. Proc. Landf.* **2014**, *39*, 1413–1420. [CrossRef]
28. Agarwal, S.; Snavely, N.; Simon, I.; Seitz, S.M.; Szeliski, R. Building Rome in a day. In Proceedings of the 2009 IEEE 12th International Conference on Computer Vision, Kyoto, Japan, 27 September–4 October 2009; pp. 72–79.
29. Wu, C. Critical configurations for radial distortion self-calibration. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 28 June 2014; pp. 25–32.
30. Cucci, D.A.; Rehak, M.; Skaloud, J. Bundle adjustment with raw inertial observations in UAV applications. *ISPRS J. Photogramm. Remote Sens.* **2017**, *130*, 1–12. [CrossRef]
31. Jiang, S.; Jiang, W.; Huang, W.; Yang, L. UAV-Based Oblique Photogrammetry for Outdoor Data Acquisition and Offsite Visual Inspection of Transmission Line. *Remote Sens.* **2017**, *9*, 278. [CrossRef]
32. Nex, F.; Remondino, F. UAV for 3D mapping applications: A review. *Appl. Geomat.* **2014**, *6*, 1–15. [CrossRef]
33. Przybilla, H.J.; Bäumker, M.; Luhmann, T.; Hastedt, H.; Eilers, M. Interaction between direct georeferencing, control point configuration and camera self-calibration for RTK-based UAV photogrammetry. *ISPRS-Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2020**, *XLIII-B1-2020*, 485–492. [CrossRef]

34. Unger, J.; Reich, M.; Heipke, C. UAV-based photogrammetry: Monitoring of a building zone. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.-ISPRS Arch.* **2014**, *40*, 601. [CrossRef]
35. Nex, F.; Armenakis, C.; Cramer, M.; Cucci, D.A.; Gerke, M.; Honkavaara, E.; Kukko, A.; Persello, C.; Skaloud, J. UAV in the advent of the twenties: Where we stand and what is next. *ISPRS J. Photogramm. Remote Sens.* **2022**, *184*, 215–242. [CrossRef]
36. Wackrow, R.; Chandler, J.H. Minimising systematic error surfaces in digital elevation models using oblique convergent imagery. *Photogramm. Rec.* **2011**, *26*, 16–31. [CrossRef]
37. Elhakim, S.F.; Beraldin, J.A.; Blais, F. Critical Factors and Configurations for Practical 3D Image-Based Modeling. In Proceedings of the 6th Conference on 3D Measurement Techniques, Zurich, Switzerland, 22–25 September 2003; pp. 159–167.
38. Nesbit, P.; Hugenholtz, C. Enhancing UAV–SfM 3D Modtheel Accuracy in High-Relief Landscapes by Incorporating Oblique Images. *Remote Sens.* **2019**, *11*, 239. [CrossRef]
39. Farella, E.M.; Torresani, A.; Remondino, F. Refining the Joint 3D Processing of Terrestrial and UAV Images Using Quality Measures. *Remote Sens.* **2020**, *12*, 2873. [CrossRef]
40. Jacobsen, K. Geometric handling of large size digital airborne frame camera images. *Opt. 3D Meas. Tech. VIII Zürich* **2007**, *2007*, 164–171.
41. Tang, R.; Fritsch, D.; Cramer, M. New rigorous and flexible Fourier self-calibration models for airborne camera calibration. *ISPRS J. Photogramm. Remote Sens.* **2012**, *71*, 76–85. [CrossRef]
42. Ebner, H. Self calibrating block adjustment. *Bildmess. Und Luftbildwessen* **1976**, *44*, 128–139.
43. Bentley. *Context Capture*. 2022. Available online: https://www.bentley.com/en/products/brands/contextcapture (accessed on 10 March 2022).
44. Agisoft. *Agisoft Metashape*. 2022. Available online: https://www.agisoft.com (accessed on 10 March 2022).
45. DJI. DJI Terra. 2022. Available online: https://www.dji.com/cn/dji-terra?site=brandsite&from=nav (accessed on 10 March 2022).
46. Jacobsen, K.; Cramer, M.; Ladstädter, R.; Ressl, C.; Spreckels, V. DGPF-project: Evaluation of digital photogrammetric camera systems geometric performance. *Photogramm. Fernerkund. Geoinf.* **2010**, *2*, 83–97. [CrossRef]
47. Nex, F.; Gerke, M.; Remondino, F.; Przybilla, H.J.; Bäumker, M.; Zurhorst, A. ISPRS Benchmark for Multi-Platform Photogrammetry. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2015**, *II-3/W4*, 135–142. [CrossRef]
48. Lowe, D.G. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]