Article

Tridimensional Reconstruction Applied to Cultural Heritage with the Use of Camera-Equipped UAV and Terrestrial Laser Scanner

Zhihua Xu [1], Lixin Wu [2,*], Yonglin Shen [3], Fashuai Li [4], Qiuling Wang [1] and Ran Wang [4]

[1] Key Laboratory of Environment Change and Natural Disaster, Ministry of Education, Beijing Normal University, Beijing 100875, China; E-Mails: hzxurs@mail.bnu.edu.cn (Z.X.); wangqiuling_2008@126.com (Q.W.)

[2] Center for Internet of Things, China University of Mining and Technology, Xuzhou 221116, China

[3] Faculty of Information Engineering, China University of Geosciences, Wuhan 430074, China; E-Mail: yonglinshen@gmail.com

[4] College of Geoscience and Surveying Engineering, China University of Mining and Technology (Beijing), Beijing 100083, China; E-Mails: lifashuai@126.com (F.L.); wangr0225@163.com (R.W.)

* Author to whom correspondence should be addressed; E-Mail: awulixin@263.net; Tel./Fax: +86-10-5880-6340.

Abstract: No single sensor can acquire complete information by applying one or several multi-surveys to cultural object reconstruction. For instance, a terrestrial laser scanner (TLS) usually obtains information on building facades, whereas aerial photogrammetry is capable of providing the perspective for building roofs. In this study, a camera-equipped unmanned aerial vehicle system (UAV) and a TLS were used in an integrated design to capture 3D point clouds and thus facilitate the acquisition of whole information on an object of interest for cultural heritage. A camera network is proposed to modify the image-based 3D reconstruction or structure from motion (SfM) method by taking full advantage of the flight control data acquired by the UAV platform. The camera network improves SfM performances in terms of image matching efficiency and the reduction of mismatches. Thus, this camera network modified SfM is employed to process the

overlapping UAV image sets and to recover the scene geometry. The SfM output covers most information on building roofs, but has sparse resolution. The dense multi-view 3D reconstruction algorithm is then applied to improve in-depth detail. The two groups of point clouds from image reconstruction and TLS scanning are registered from coarse to fine with the use of an iterative method. This methodology has been tested on one historical monument in Fujian Province, China. Results show a final point cloud with complete coverage and in-depth details. Moreover, findings demonstrate that these two platforms, which integrate the scanning principle and image reconstruction methods, can supplement each other in terms of coverage, sensing resolution, and model accuracy to create high-quality 3D recordings and presentations.

1. Introduction

The use of a 3D model derived from a cultural or historical heritage domain has recently become the focus of considerable attention for different purposes, such as preservation [1], as-built documentation [2], reconstruction [3], and museum exhibitions [4]. This model has also become an important component of the modern digital age. In fact, a significant number of initiatives in the form of research and development projects have focused on establishing 3D documentation as an affordable, practical, and effective mechanism that enables the content enrichment of cultural or historical heritage digital libraries [5–7]. The 3D reconstruction method has recently become an important procedure, particularly for historical heritage preservation capability. This procedure was added to the UNESCO World Heritage list in 1985, thus proving that 3D reconstruction is vital to the conservation and preservation of building complexes.

Most available instruments or techniques produce high-quality results for 3D reconstruction; these techniques are thus accepted as common practices for recording historical or cultural heritage objects [8]. The most common techniques are based on classical surveying (e.g., total station, Global Navigation Satellite System (GNSS)/Global Positioning System (GPS)) and/or traditional photogrammetry with control points and a human operator [1]. This task is time consuming and requires sustained attention. However, 3D reconstruction techniques that rely on laser scanning and a large number of automated image-based methods have recently become available [3].

TLS systems have been widely used for cultural or historical heritage documentation, as well as for 3D reconstruction and visualization [2]. The scanner measures the distance from the scanner to the object surface either by using the time-of-flight (TOF) principle or by measuring the waveform between transmitted and received signals [2]. In the TOF principle, distance is computed on the basis of the measurement of the response time of a light beam between transmission and reception. Meanwhile, in the second method, distance is computed by comparing the wave difference between the transmitted and received signals [9,10]. With the known triangulation principle, this process will

produce a group of 3D points. Each point is indexed with a coordinate position as well as with the intensity of the returned pulse.

Numerous researchers have reconstructed 3D models that have been applied to cultural heritage objects with the use of a TLS device, which directly produces 3D points and provides dense geometry information. In addition, TLS clouds produce a 3D model for an object with realistic size and high accuracy, even though some range scanning techniques result in a cloud with rough accuracy.

Image-based 3D reconstruction techniques are considered cost-effective and efficient for producing a high-quality 3D digital model of real-word objects in terms of hardware requirements, knowledge background, and man-hours [11]. At least two images with common features are generally required. Additionally, 3D data accompanied by texture information can be derived through perspective or projective geometry formulations [12]. These methods (mainly computer vision [13]) are generally preferred in case of lost objects, monuments, or architectures with regular or complex geometric shapes, small objects with free-form shape, mapping applications, deformation analysis, and time or location constraints for data acquisition and processing [1]. The most popular among these available techniques are the structure from motion (SfM) [14] and dense multi-view 3D reconstruction (DMVR) [15] algorithms. These methods can recover scene geometry by processing a number of unordered images that depict a scene from perspective viewpoints with high overlaps. The pipeline is generally composed of different steps that are automatically performed. First, camera parameters are automatically computed by matching the corresponding features shared between different views that depict overlapping areas [14]. The bundle adjustment method is typically employed to improve the accuracy of camera trajectory calculation, minimize projection error, and prevent the error buildup of camera tracking [16]. The SfM technique attempts to recover camera parameters and a sparse point cloud of the scene, whereas the DMVR method focuses on improving the thoroughness of detail in the point cloud. A number of software solutions that integrate these algorithms have been made available to the general public over the last decades. Anestis *et al.* [11] provided a detailed review of state-of-the-art image-based 3D reconstruction techniques involved in research, projects, or systems. For instance, the open source Bundle [17] is accepted as the basis of the SfM system and has thus undergone successive improvements [18–20]. Moreover, Patch-based Multi-view Stereo (PMVS) [15] is accepted as one of the most popular DMVR algorithms. A large number of commercial software programs, such as Visual SFM [21], PhotoModeler [22], and PhotoScan [23], that integrate SfM and DMVR algorithms are available. Moreover, with the emergence of UAV and its applications, Pix4D developed the Pix4UAV [24] software that aims at processing UAV optical sequences to create 3D digital elevation models. The camera-equipped UAV platform overcomes the problems of traditional photogrammetry in terms of launching site constraints, camera calibration, and high cost. Furthermore, the UAV platform, especially rotors with hovering capability, enables the possibility of area measurements, which poses location constraints in TLS scanning. This feature makes up for the incapability of TLS to measure building roofs. Despite these successes in the 3D reconstruction of cultural heritage objects, numerous challenges remain. Remondino *et al.* [1] highlighted the actual problems and main challenges confronted in every process of the 3D reconstruction of large and complex sites or objects. These processes ranged from data acquisition to the visualization of the final 3D models. A significant challenge is posed by the selection of the appropriate platform for data acquisition, data processing, data quality, and complexity, as well as geometric resolution variety.

For TLS measurement, one limitation is attributed to the additional high cost entailed by such device. Moreover, geometry can be extracted with unusable texture information because such systems are equipped with an inner camera, which is sensitive to illumination. Additionally, the procedures for terrestrial surveying typically capture information on facades and are beyond the range of measuring building roofs unless the device is placed higher than the object, which further limits the universal application of this method. For SfM 3D reconstruction, the main problem is the time consumed by the added image, which requires a search for all candidates during image matching and thus results in significant redundancy [25,26]. Moreover, traversal matching also results in a large number of mismatches between pairs of images without common features. Furthermore, model accuracy is unstable and is affected by the quality of sequence initialization and bundle adjustment. In most cases, the model errors reportedly reach 5% (accuracy of *c.* 1:20); however, these errors can be improved to an order of magnitude through a carefully planned network [27].

The combination or integration of hybrid sensors and techniques has been proven to be the best solution for 3D modeling applied to heritage sites with large or complex structures [28]. A large number of previous studies reported on the integration of laser scanning and digital photogrammetry, such as surveying and 3D modeling techniques, for application to heritage sites (Linguaand Rinaudo [29]; Ioannidis and Tsakiri [30]; Beraldin [31]; Boehler and Marbs [32]; Kadobayashi *et al.* [33]; Remondino *et al.* [34]). However, integration is mainly exploited at the model level in most cases, despite the need to conduct such process at the data level to overcome the weakness of each data source [1].

In this study, we integrate multi-source clouds at the data level. We focus on the designation of improvements in 3D reconstruction applied to cultural heritage in terms of data acquisition, data processing efficiency, model quality, and accessibility. In particular, we discuss the idea of integrating laser scanning and image-based 3D reconstruction techniques to produce complete and detailed 3D models for objects of interest relative to heritage objects. In addition, we compute a camera network by analyzing the image conjunction according to the rough flight control data to improve SfM performance in terms of matching efficiency as well as to reduce false matches. Moreover, we assess the applicability of the integration between TLS and UAV platforms for measuring complete 3D points with an iterative registration method, from coarse to fine. To exploit the intrinsic advantages of each sensor, UAV data may be used to reconstruct the basic shape and main structural elements through image reconstruction techniques. Meanwhile, laser scanning data may be used for fine details and sculpted surfaces.

The remainder of this paper is organized as follows: Section 2 presents a detailed description on the methodology of the complete pipeline. Section 3 introduces the experimental study and the procedures used for data collection. Section 4 evaluates the performance of the proposed methods and discusses the results in terms of scene geometry 3D reconstruction and modeling. Section 5 concludes.

2. Methodology

This section introduces the methodology employed to construct complete 3D models of large and complex objects of interesting for heritage objects. Figure 1 is adopted from [35] and presents

the frame of 3D reconstruction of cultural heritage that integrates a camera-equipped UAV platform and TLS. The approach mainly consists of two aspects:

1. Recover the scene geometry by orienting and calibrating the bundle of UAV images with the use of our modified SfM method with a computed camera network; and
2. Coarse-to-fine co-registration of multi-source point clouds with an iterative method.

Figure 1. General pipeline for delivering photorealistic 3D models from camera-equipped unmanned aerial vehicle system (UAV) platform and terrestrial laser scanner (TLS).
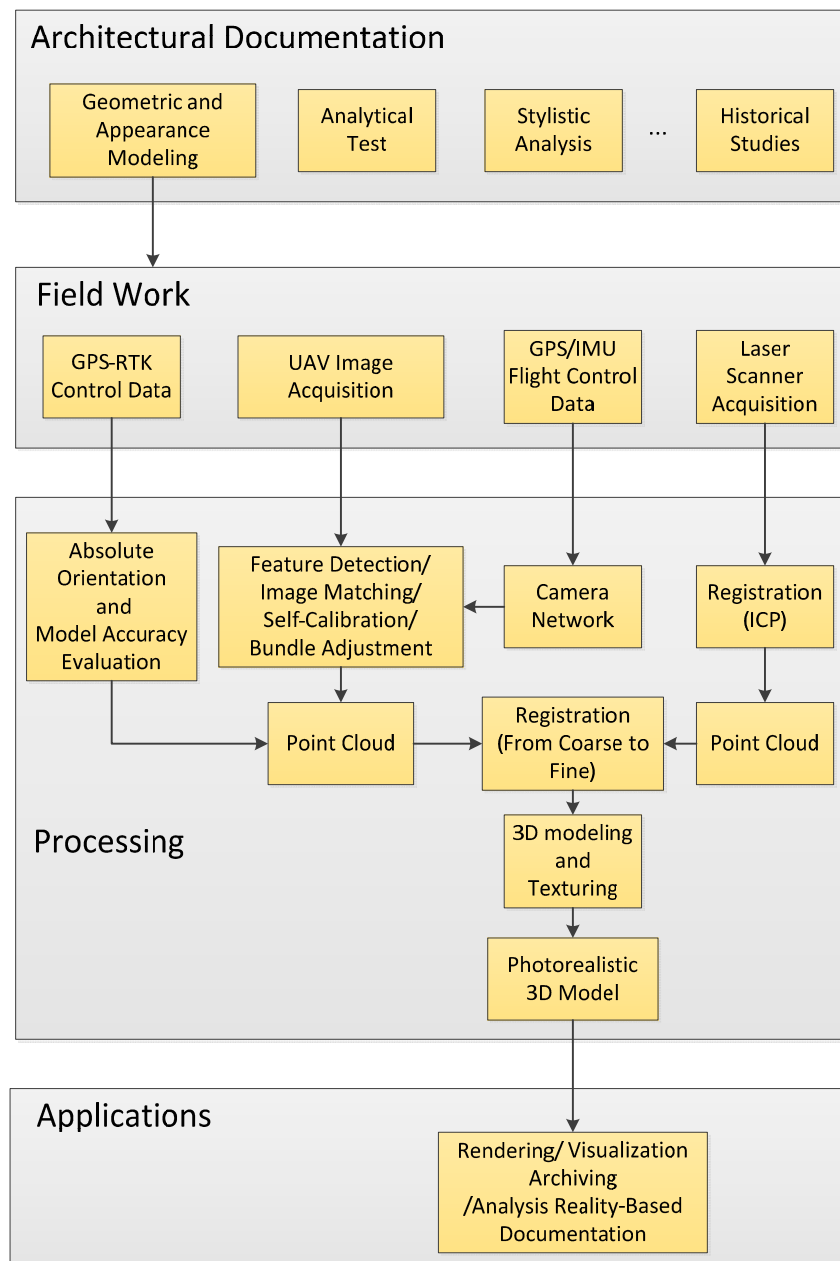

Additional flight control information on the airborne GPS and IMU devices is used to compute for a topological camera network to guide the image matches. The control points are used to address absolute orientation after processing the geodetic data acquired by the GPS-RTK. These points are also used to evaluate the accuracy of the point cloud particularly that extracted from UAV optical collections.

2.1. Image-Based Reconstruction of the UAV Optical Images

The widely used SfM is implemented as a cost-effective method for processing sets of overlapping UAV collections in the point cloud generation procedure. SfM uses pre-detected features to create a sparse point cloud. Figure 2 describes the SfM pipeline for perspective collections. This pipeline mainly contains procedures for distinctive feature point detection, image matching, and bundle adjustment. These procedures aid in establishing the scene geometry and camera parameters. The focal length of the camera should be fixed when adapting the steps employed in SfM to simplify bundle adjustment and self-calibration. The input images should also have a large degree of overlap with each part of the surface to be reconstructed that is visible in at least three images [36]. We utilize the applicability of this pipeline for UAV optical collections and propose necessary improvements or modifications for different steps as needed.

Figure 2. Pipeline of the applied camera network modified structure from motion (SfM) for processing UAV optical collections.
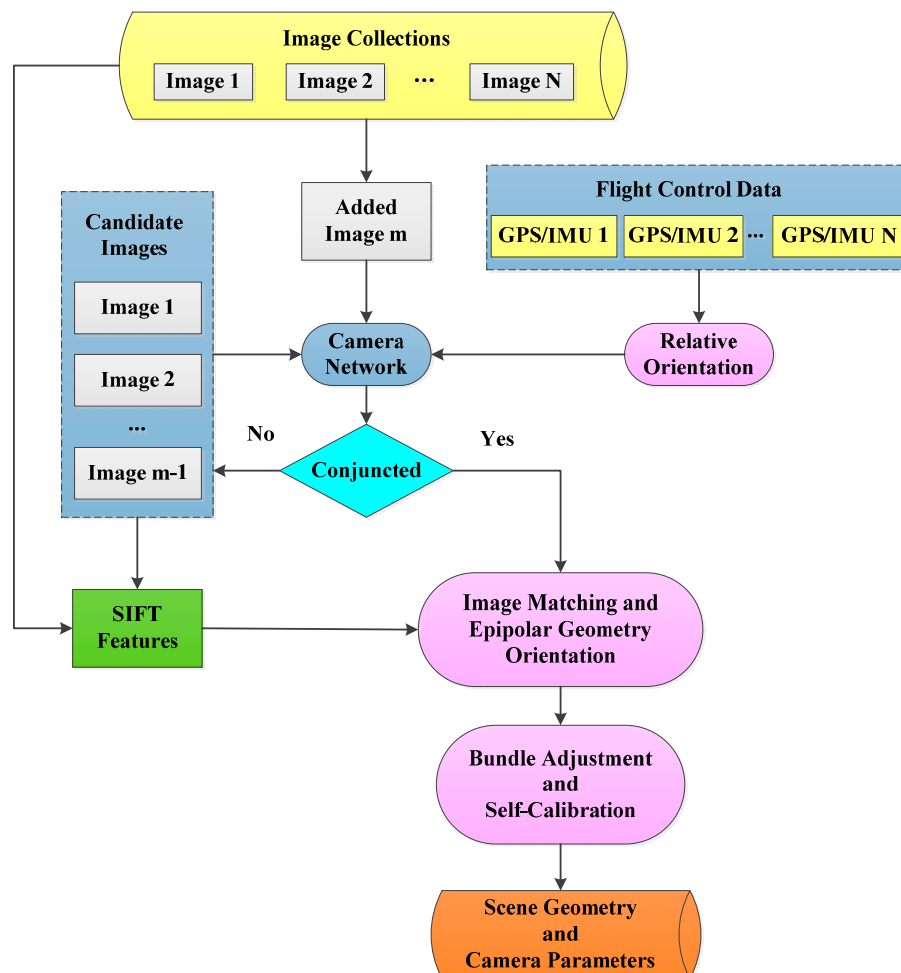
The first step is to load the images for the analysis and automatic extraction and matching the correlated feature points in multiple images. The well-known Scale-Invariant Feature Transform (SIFT) [37] algorithm in Bundler is one of the most excellent descriptors for its invariance to image scaling, camera rotation, and a certain degree of illumination. Thus, SIFT is used to detect and match

features. However, based on our experiment experience, SIFT usually results in memory explosion when dealing with UAV images with very high resolution. Hence, we divide the image into blocks of equal size, extract the local extreme points in each block, and then combine the subset of the feature points.

For image matching, the added image should successively search for all candidate images and identify the correlated feature points between the image pair with a threshold that evaluates the Euclidean distance similarity [25]. Consequently, the number of computations required to match the feature points is quadratic in terms of the number of images, which results in significant redundancy [38]. Notably, threshold-based traversal image matching also results in significant mismatches between pairs of images without common features. Therefore, we propose a camera network by taking full advantage of the airborne flight control data to constrain the matching track to reduce computation cost and drop the mismatches. The camera network is computed through image topological analysis.

First, all images are projected onto the user coordinate through relative orientation with reference to the airborne flight control information. Meanwhile, we compute for the positions of vertexes for each UAV collection by using photogrammetry equations [26].

Second, we compute the topological relationship between one image P and each vertex V in the other image as $T(P,V)$.

$$T(P,V) = \begin{cases} 1, & S_P = S_{\Delta V_{P12}V} + S_{\Delta V_{P23}V} + S_{\Delta V_{P34}V} + S_{\Delta V_{P41}V} \\ 0, & S_P \neq S_{\Delta V_{P12}V} + S_{\Delta V_{P23}V} + S_{\Delta V_{P34}V} + S_{\Delta V_{P41}V} \end{cases} \tag{1}$$

where $T(P,V) = 1$ indicates that the vertex V is contained in image P, whereas $T(P,V) = 0$ indicates that vertex V lies outside of image P. S_p represents the area of image P, and the formula to the right of each item indicates a triangle area composed of the vertex V and any two sequential vertexes of image P.

Finally, we identify the conjunction relationship between nearby images by determining whether any vertex from one image lies in the coverage of the other image. If none of the vertexes from one image lie in the coverage of the other, no common features are found between this image pair. Thus, the image topology between image P_l and P_k is marked as $T(P_l, P_k) = 0$; otherwise, $T(P_l, P_k) = 1$.

$$T(P_k, P_l) = \begin{cases} 1, & \sum_{j=1}^{4} T(P_k, V_{lj}) + \sum_{j=1}^{4} T(P_l, V_{kj}) \neq 0 \\ 0, & \sum_{j=1}^{4} T(P_k, V_{lj}) + \sum_{j=1}^{4} T(P_l, V_{kj}) = 0 \end{cases} \tag{2}$$

To simplify, we consider the UAV collections as network $G(V, E)$, which has a node for each image and a directed edge between any image pair with common features.

$$V(G) = \{P_1, P_2, P_3, ..., P_n\} \tag{3}$$

$$E(G) = \{(P_i, P_j) \mid T(P_i, P_j) = 1; P_i, P_j \in V(G)\} \tag{4}$$

where n indicates the amount of UAV collections.

In this way, the added image will search for corresponding images on the basis of the computed camera network $G(V,E)$. Feature points will not be matched if such points are not present in a pair of images with common areas. Consequently, the probabilities of mismatching the features between

an image pair without overlaps decrease, and SIFT features produce a significantly direct matching. For more details on network construction, please refer to our previous work in [26].

In the following step, a bundle adjustment operator in Bundle is implemented by minimizing the distances between projection rays and feature matches. This process results in a 3D point cloud. Each point contains RGB color information extracted from the correlated multiple images. The density of the point cloud is sparse and can be densified to a dozen orders of magnitude with PMVS [15].

2.2. Co-Registration of UAV Image-Reconstructed Cloud and TLS Scanning Cloud

The co-registration process for the two groups of point clouds finds the Euclidean motion from one to the other, such that all point clouds are represented with respect to a common coordinate system. Recently, the well-known registration methods include Iterative Closest Point (ICP) algorithm [39] and its variants. The basic idea of ICP is to treat the nearest point in the other cloud as a correlated point. Given that image-based 3D reconstruction and laser scanning have different principles, coordinates of the two groups of clouds are independent of each other. Consequently, the point clouds extracted from the UAV image reconstruction and TLS measurement present differences in terms of sensing resolution, coverage, accuracy, and scale. Directly implementing an ICP operator will result in a fatal error. Thus, prior coarse registration is necessary for a successful ICP. Another issue that should be considered is the sufficient overlap between the two groups of point clouds.

TLS generally measures the building surface from multiple surrounding stations to create multiple scans that mainly cover information on side walls. Building roofs cannot be covered unless the TLS is placed higher than the objects. An eight-rotor UAV platform is used as supplement. This device acquires information beyond the TLS measurement range. The point cloud that—possesses higher resolution and model accuracy is selected as reference in the co-registration procedure. Before registration, partial scans from the TLS are registered in advance with the ICP module, which is integrated in the software RiSCAN PRO.

In coarse registration, the two groups of point clouds from UAV image reconstruction and TLS measurement are separately oriented to the user coordinate by geodetic ground control points (GCPs) with the well-known Bursa transformation model. Consequently, two clouds with the same scale are produced. A set of correlated features from both clouds lies in the roof boundaries, corners, and road crossings are manually selected to compute for the initial motion between the two clouds. The result of coarse registration is generally poor. Thus, fine registration is subsequently conducted to improve data merging accuracy. The critical issue involved in fine registration is finding the corresponding features between the two groups of 3D points. With an ICP operator, correspondences are established by searching for points in the destination cloud that are closest to a set of points in the source cloud. The nearest point matches are not reasonable correspondences because the sensing resolution is different between the two clouds. Thus, the point-to-point minimization criterion will introduce a number of mismatches. To enhance performance, the normal constraint is considered to reject incorrect correspondences according to one of the ICP variants: Iterative Closest Compatible Point (ICCP) [40,41]. Finally, the Euclidean motion between two point clouds is iteratively computed by minimizing the cost function C with a least square method, as presented by Formula (5). All

the correspondences are used for the adjustment. A weight of 1.0 is recommended for point coordinates from both image reconstruction and TLS measurement.

$$C = \arg\min \left\| \sum_{i=1}^{m} \left(x_1' - x_0 \right)^2 + \left(y_1' - y_0 \right)^2 + \left(z_1' - z_0 \right)^2 \right\| \tag{5}$$

$$\begin{bmatrix} x_1' & y_1' & z_1' \end{bmatrix} = kR \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} + T \tag{6}$$

where (x_1, y_1, z_1) and (x_0, y_0, z_0) represent the points in the image-level and TLS clouds, respectively; and k, T, and R represent the Euclidean motion parameters for a constant scale, translation column, and a 3 × 3 matrix, respectively.

Although digitization and modeling are not the main focus of this study, with the final 3D points, such further steps, as digitalization, modeling, and texturing are implemented to obtain a high-resolution photorealistic 3D model. The segmentation and structuring phase is conducted before producing a surface model. Additionally, outlines with simple elements, such as block, circle, or other polygonal boxes, are extracted for model digitalization. Finally, airborne or terrestrial images taken from perspective viewpoints are projected onto the 3D facades through image-to-geometry registration. This projection results in multiple partial facades, which aid in the establishment of a complete textured model.

3. Experimental Study

In this section, we provide some historical information on the heritage object. We also describe the data collection procedures and equipment used.

The experimentation yielded interesting results with respect to depth resolution, model completeness, and accuracy. To validate the reconstruction procedure on real objects, the methodology has been applied to a historical monument. The complexity of this monument enabled the performance evaluation of the combination of UAV image-reconstructed cloud and TLS cloud for 3D reconstruction. The accuracy of the two groups of point clouds is independently evaluated.

3.1. Study Area

The methodology can be applied to objects, especially those with large size and complex construction, for which no single sensor can capture all information. Thus, we chose a case study that covers a historical monument with complex shapes. The monument is called Liao Family Temple, which was built at the end of the Qing Dynasty (1644–1911). Liao Family Temple is at the foot of Bijia Mountain, Gutian Village in Fujian Province, China and covers approximately 826 squares. The temple is a group of brick-and-wood structures consisting of the front and back halls and the wing rooms. This structure has enamel detail and a series of roof corners. Liao Family Temple is considered as one of the most important monuments both in culture and in politics of this area. The great civil war memorial dedicated to this monument has prompted its acceptance as the Gutian Conference Site of the Red Army of Chinese Communist Party in 1929. This monument has also suffered from the

aftershocks of successive historical events, which ranged from the Second World War to present-day incidents. Hence, Liao Family Temple is in dire need of reconstruction for preservation.

3.2. Data Acquisition

The fieldwork was separated into two sessions. The first session involved aerial imaging. The second session involved terrestrial laser scanning, followed by empirical measurements and terrestrial imaging.

3.2.1. UAV Oblique Image Acquisition

An eight-rotor UAV platform called BNU D8-1 is employed for the aerial image shooting session (Figure 3a). This UAV platform is equipped with a three-axis pan-tilt-roll remote-controlled camera head (360° on the horizontal axis, 110° on the vertical axis, and rolling ability of 60°). A DSLR Canon EOS 5D Mark II at 8.1 MP with an 18–55 mm lens is used for the aerial imaging session. Image resolution at a typical flying height above terrain of 100 m is approximately 2 cm/pixel. The BNU-D8-1 has a payload limit of approximately 5 kg and full payload flight duration of approximately 20 min. BNU-D8-1 has an onboard navigation system with a navigation-grade GPS receiver (U-bloxLEA6H) and three small MEMS-based IMUs. This system obtains camera positions and postures at the time of camera exposure. BNU-D8-1 can be controlled manually or automatically through self-learning neural network adaptive control technology (Brainy BEE autopilot system V4.11). In this study, BNY-D8-1 is at an automatic route with controlled height above the terrain of 100 m. The route contains three regular lines, thus that we obtain 45 images that possess 80% endlap and 60% sidelap.

Figure 3. Equipment used for data acquisition. (**a**) Eight-rotor UAV platform (BNU-D8-1) fitted with Cannon 5DII; (**b**) Riegl VZ-4000 TLS.

(**a**)   (**b**)

3.2.2. Terrestrial Surveying

In our case study, the short range of the terrestrial laser scanner (TLS) is unavailable. Thus, we used Riegl VZ-4000 (Figure 3b), which is capable of extremely long-range measurement, to capture the surface information of the Gutian Conference Site. The system specifications indicate a minimum distance of 5 m between the scanner and the surface to be scanned. The system offers 15 mm standard deviation error for measurements implemented at a 150 m distance and a 1 cm maximum distance between two sequential points within a 1000 m distance. The integrated digital camera of the scanner

has a 5 MP CCD sensor. This camera is sensitive to illumination, which may result in poor color quality.

Six partial scans were captured from multiple stations. Subsequently, the multi-scan registration procedure was conducted with the use of the standard procedure implemented by manual coarse registration and multi-station-adjustment modules in the RiSCAN PRO 1.1.7 software package. The average distance from the monument was estimated at 30 m, whereas the average distance between two consecutive points was approximately 2 cm. Moreover, the density of the scanning cloud was 2500 points/m^2.

Additionally, 16 GCPs were evenly distributed around the monument. These GCPs were pre-measured by using a GPS-RTK instrument, resulting in a maximum of ±1 cm positioning error. These GCPs can be used in two ways, *i.e.*, for model orientation and for assessing model accuracy.

A consumer digital Canon 50D at 15.1 MP with an 18–55 mm lens was used for the terrestrial imaging session. The average distance of the camera from the monument surface was estimated as 6 m. Perspective images were captured with an average resolution of 1 cm/pixel. These images were subsequently used for model texturing.

4. Results and Discussions

This study mainly focuses on the 3D reconstruction of the outdoor surface for Gutian Conference Site. Nevertheless, the proposed approach is valid for partial measurement with the use of the TLS or rotor UAV platform. The performance of the camera network matching method is discussed. The complete 3D points are presented by combining the UAV image-reconstructed and TLS clouds. Finally, the final photorealistic 3D model of the monument is established.

4.1. Point Cloud Generation from UAV Optical Images

A point cloud is computed with the use of proposed camera network matching method by processing a set of UAV optical images. In this way, feature points will not be matched if they are not in the image pair with a direct edge. Consequently, the camera network revealed that considerable traversal matching redundancy and mismatches are eliminated in the image matching and image orientation processes.

Figure 4a presents the flight information that contains three lines and indicates the conjunction between pairs of nearby lines. Figure 4b,c illustrates the performances of match features in a traversal manner and with the proposed camera network, respectively. We construct an $n \times n$ (n is the image number, which is 45 in this study) grid and project the match numbers between pairs of images into the corresponding grids. The match numbers on the grids are proportionally visualized with color degree. When more correspondences exist between image pair, the map pixel shows a higher level of color degree. The results show several matches between the nearby images in the heading direction and between images with side laps. Notably, Figure 4b displays a great deal of noise with low color intensity between pairs of images without overlaps. These noises indicate the mismatches, and are well addressed by our proposed camera network matching, as shown in Figure 4c.

Figure 4. Comparison of matching performance between traversal matching and camera network guided matching methods. (**a**) Sketch map represents the flight information; (**b**) Matched features between any pair of images in the traversal way; (**c**) Matched features between any pair of images with the camera network guided method.
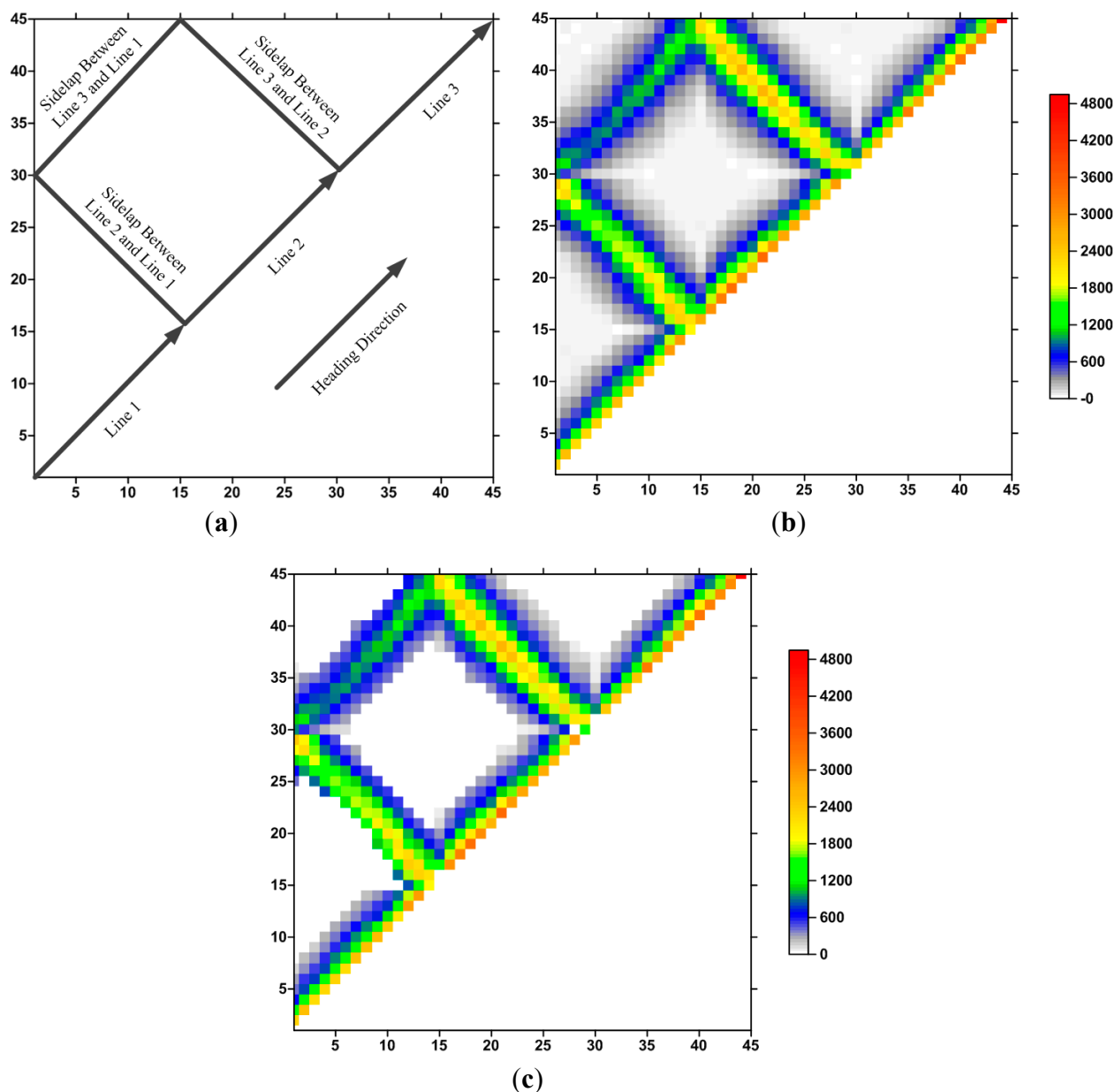

(a)

(b)

(c)

Table 1 shows the results of the automated orientation in terms of time consumed and number of reconstructed tie points for both the traversal matching and camera network guided methods. To simplify, we record the matching cost with the number of matching image pairs. Traversal matching produces a total 900 matching image pairs, indicating an $O(n^2)$ computation complexity for 45 input images. The camera network guided matching method introduces a total of 564 matching image pairs. The efficiency of the camera network is nearly twice as high as that of the traversal matching strategy. Nevertheless, the camera network achieves performances comparable with those of traversal matching in terms of sparse points and the dense points regardless of the large reduction in the number of matching images.

Table 1. Performance comparison in terms of matching cost and number of output points between traversal matching and camera network guided methods.

	Matching Images	**Sparse Points Number**	**Dense Points Number**
Traversal Method	900	3.02×10^4	3.21×10^6
Camera Network	564	2.98×10^4	3.21×10^6

Figure 5 presents the perspective views of the sparse and dense point clouds that were reconstructed with the camera network matching method. A remarkable increase in point density is immediately apparent for the dense reconstruction. After manual editing, the sparse dataset comprised 3.02×10^4 points with a density of 4 points/m^2, whereas dense reconstruction produced 3.21×10^6 points with a density of 40 points/m^2, which is an order of magnitude increase. Figure 5b displays the camera network, which describes a conjunction graph for the image pairs with common features. Visualization in Figure 5c shows that the resolution is sufficient to reveal the roof structure, particularly the shape, texture, and tilling array.

Figure 5. UAV image-reconstructed point clouds. (**a**) Gutian Conference Site with distributed ground control points (GCPs); (**b**) Sparse point cloud extracted from UAV sequences matching with the camera network; (**c**) Dense point cloud.
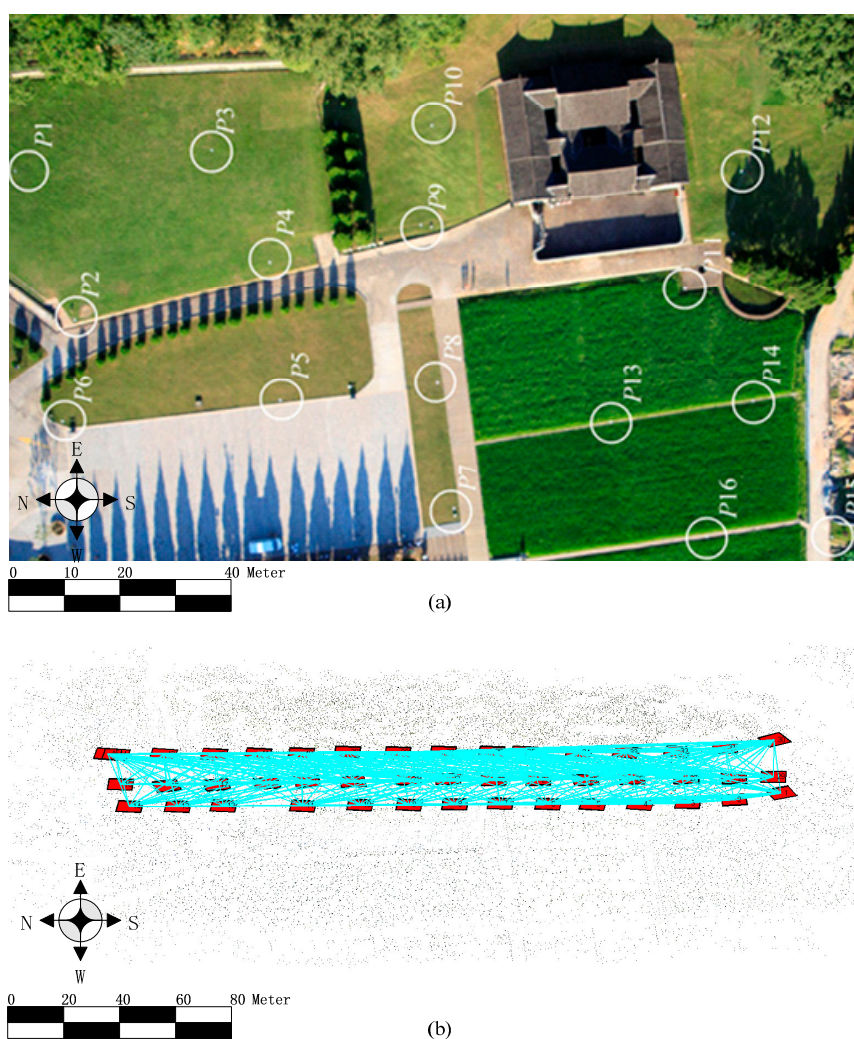

(a)

(b)

Figure 5. *Cont.*

(c)

To assess model accuracy of image-level point cloud, orientation is achieved in two steps. A relative orientation is initially applied, followed by an absolute orientation through 16 well-distributed GCPs, as shown in Figure 5a. Rather than focusing on absolute positioning, we are more concerned about the relative accuracy (ratio of difference between the measured distance and actual distance to the actual distance) between model elements. Figure 6 displays the relative error in the measurement by using eight control points and eight checkpoints. The model error was evaluated within the range of 8‰ to 12‰.

Figure 6. Model accuracy of the image-level cloud extracted from UAVs optical sequences with the camera network matching method.
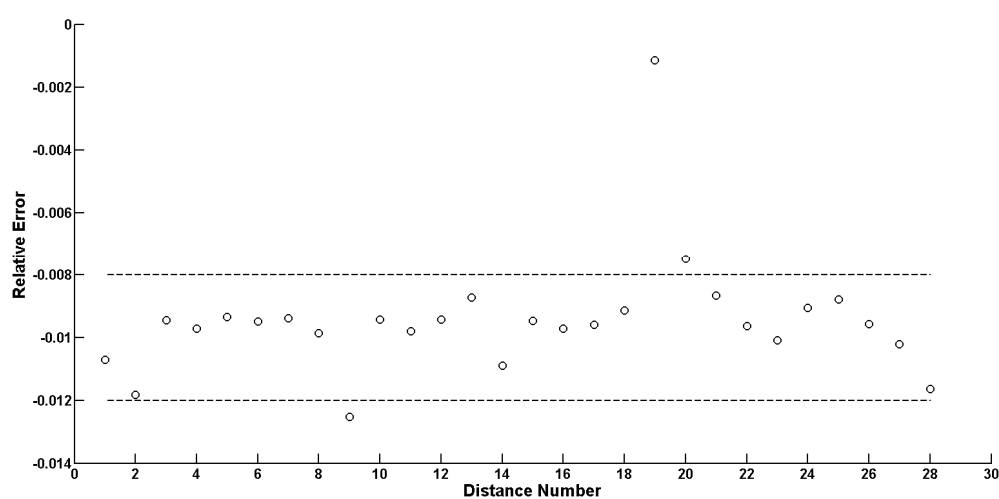

4.2. Accuracy Evaluation of TLS Point Cloud

The accuracy of a TLS cloud mainly depends on the range between the operating scanner and object surface. The scanning accuracy decreases with increasing range. We evaluate the cloud accuracy

independently in an open field with 12 distributed reflectors. All reflectors are pre-measured with a total station. The distances between the TLS instrument and the reflectors are of approximately 150 m. The results show that the absolute positioning accuracy (Table 2) is mostly consistent with the nominal accuracy, thus resulting in a standard deviation error within 15 mm of the actual value, except for a larger −19.1 mm at 158.412 m distance. The relative accuracy of TLS measurement (Figure 7) is often lower than ±1% within a 150 m range, which is an order of magnitude higher than that of image-level points.

Table 2. Absolute accuracy of the terrestrial laser scanner (TLS) measurement at approximately 150 m range indicating the degree of conformity of a measured distance to its actual value.

Reflector	1	2	3	4	5	6	7	8	9	10	11	12
Distance/m	145.371	153.80	159.593	167.760	138.470	136.468	158.412	158.353	146.910	139.477	137.431	150.075
Error/mm	−11.5	−9.0	−8.0	−10.4	−12.6	−13.3	−19.1	−10.2	−15.5	−12.7	−15.0	−14.3

Figure 7. Relative accuracy of the TLS measurement at an approximately 150 m range.
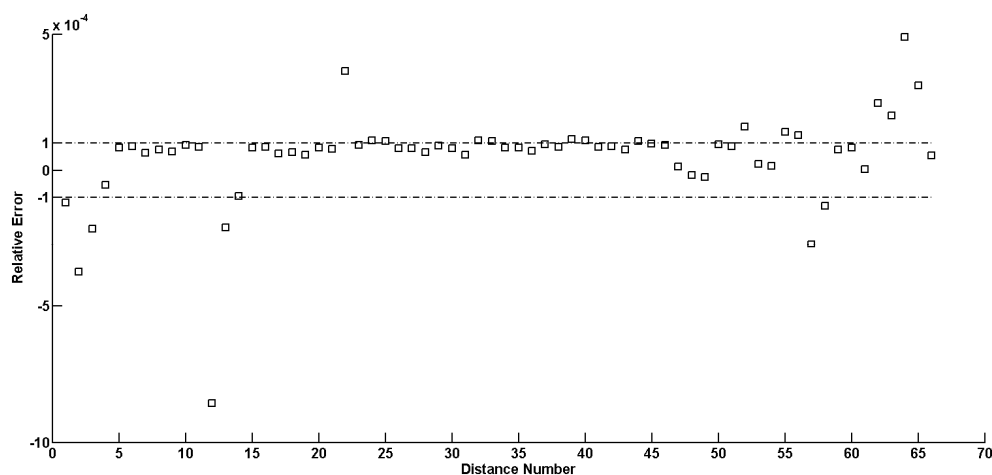
4.3. Co-Registration of Point Clouds from UAV Image Reconstruction and TLS Scanning for 3D Reconstruction

With the use of the methods illustrated in Section 2, the two groups of point clouds from UAV image reconstruction and TLS scanning are accurately registered. Given that a TLS cloud possesses higher sensing resolution than an image-level cloud with a density of 2500 points/m^2, the former is selected as the source dataset. These groups create complete 3D points. The statistical standard deviation of registration error is lower than 2.5 cm. The output of registration, *i.e.*, the multi-scans from the TLS, covers most of the information of building sides (Figure 8a), which leaves partial roof information behind (Figure 8b). Figure 8c shows that the integration of point clouds acquired by the abovementioned platforms improves the completeness of model coverage. The sensing resolution or density of the image-level cloud is evidently sparser than that of the TLS cloud. A high level of detail was unsuccessfully reconstructed in the hipped gable. Nonetheless, the details provided are sufficient to reveal the roof structure for further 3D modeling.

Figure 8. Two groups of point clouds extracted from the TLS sensor and camera-equipped UAV platform. (**a**) Result of multi-scan registration from TLS, front view; (**b**) and bird's eye view; (**c**) Final points with the integration of UAV image-reconstructed cloud and TLS scanning cloud.
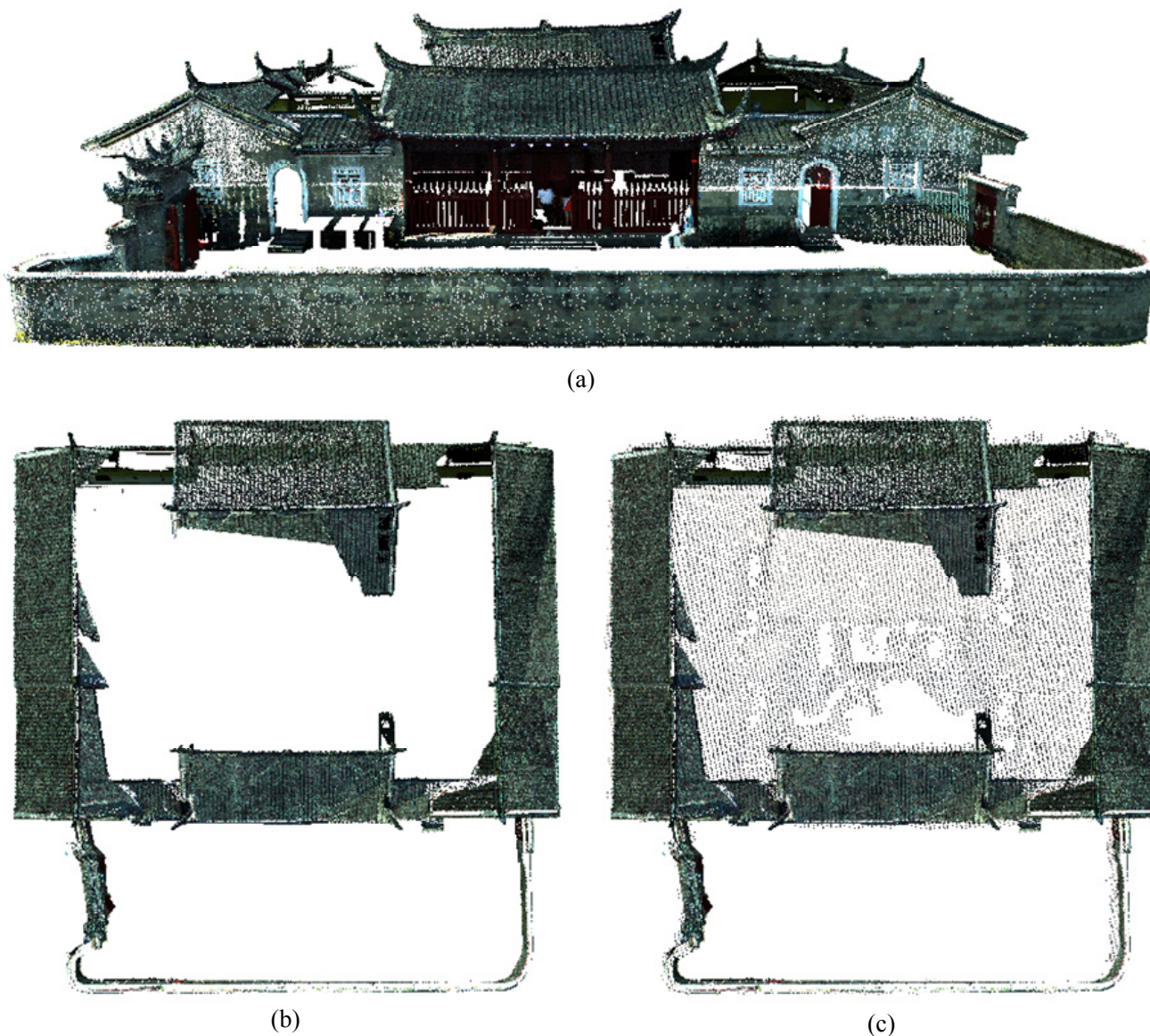

(a)

(b)

(c)

Figure 9a shows that the facades of the final 3D points are manually drawn stone by stone. Figure 9b shows the 3D model after drawing. This model covers a high level of detail in terms of eaves, goalposts, and steps. In the final stage, the photorealistic 3D model is produced by projecting the orthophotos onto the model surface, as shown in Figure 9c.

Figure 9. 3D modeling of Gutian Conference Site. (**a**) Drawings of the monument; (**b**) 3D model of the monument; (**c**) Photorealistic 3D model of the monument.

(a)

(b)

(c)
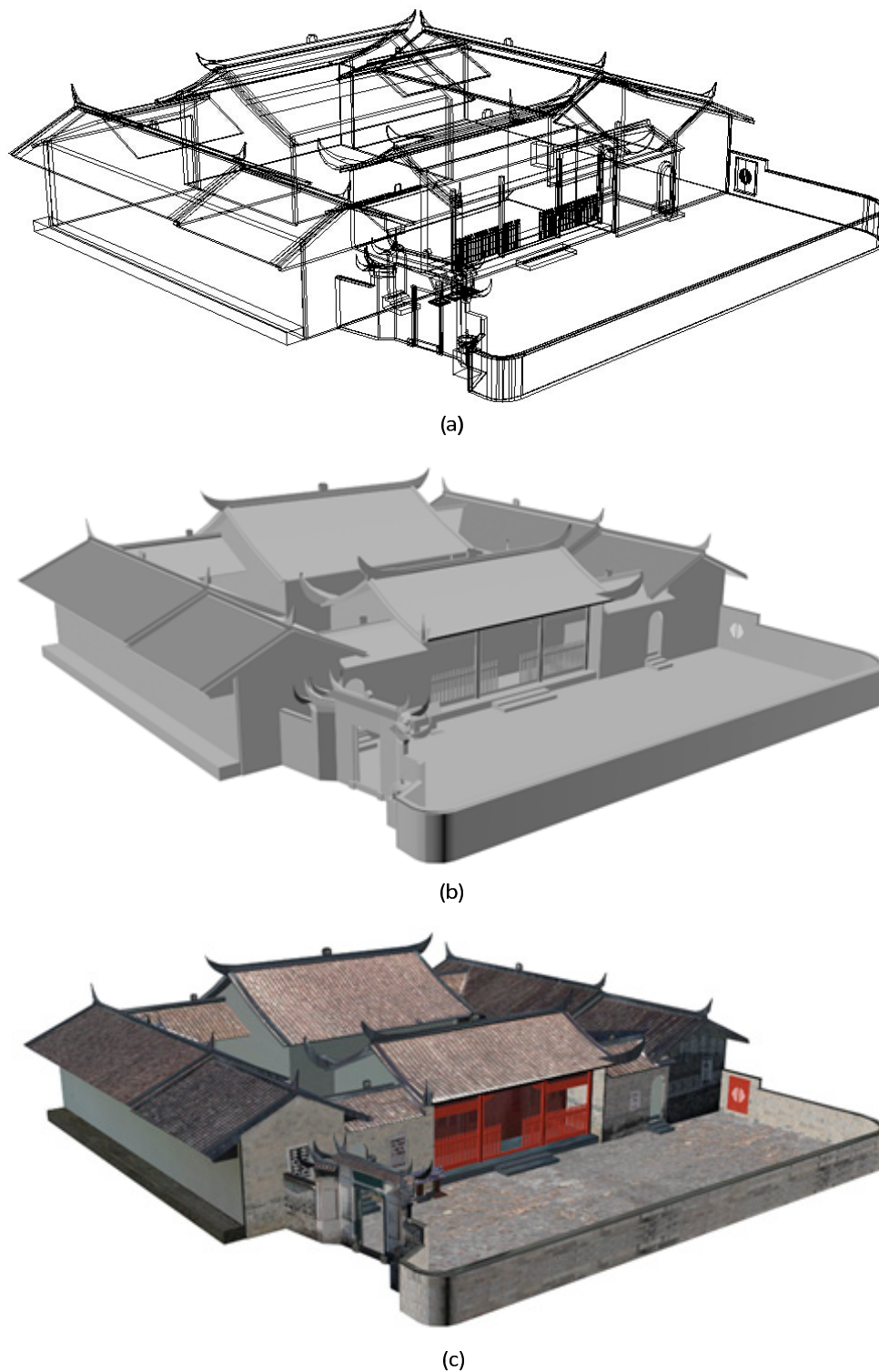
4.4. Performance and Analysis

In this study, we take the advantages of both TLS, which acquires a point cloud with high sensing resolution and model accuracy, and the camera-equipped rotor UAV platform, which captures

perspective optical sequences. Impressive results were obtained. The following issues should be considered to improve the performances significantly.

(1) The flight should cover multiple lines and should obtain images with large heading and side overlaps to enhance the stability of image orientation and geometry reconstructed results.

(2) Image resolution affects feature point extraction. Traditionally, an image with a higher resolution will produce more feature points. However, detecting SIFT or SURF features from a low-altitude UAV image generally results in memory explosion because of extremely high sensing resolution. Hence, a sub-divided strategy is introduced in [26].

(3) Airborne positioning accuracy affects the camera network. The navigation precision with the airborne GPS and IMU devices is within ±5 m, which is not as accurate as that of high-quality GNSS/INS devices. However, this accuracy is acceptable when considering the land coverage of each image (approximately 102 m × 68 m at the height of 100 m above the terrain). This precision helps identify the conjunction relationship between nearby images, which will result in insignificant false matches.

(4) Resolution difference should be considered. Point clouds captured from hybrid sensors possess different sensing resolutions and various scales. The integration of hybrid point clouds is based on the principle of feature matching. Hence, resolution difference affects the final result. The cloud, which possesses higher sensing resolution and model accuracy and was thus used as base data in the data integration procedure. Furthermore, GCPs were used to solve the scale problem.

We should primarily focus on data integration performance. One of the most significant advantages of the TLS cloud referred in our study is its high accuracy in terms of geometric measurements despite its poor texture information and high cost. Nevertheless, a camera-equipped UAV platform enables flexible data acquisition and produces a textured point cloud with slightly rough sensing resolution and model accuracy. The integration of these two sensors takes full advantage of both model accuracy and coverage completeness.

Numerous methods are suggested to improve the accuracy of integration results. One potential solution is the use of an image matching method [42,43]. In this manner, the TLS cloud is first projected onto a 2D range image and matched to UAV images on the basis of distinctive features. Subsequently, reverse projecting to the 3D coordinate is performed. Other suggestions can be implemented to improve the efficiency of the results. One suggestion is to reduce the number of images during the image matching and bundle adjustment procedures. Alsadik *et al.* [27] filtered the image network with the minimal camera principle and then applied improvements through a compromise between coverage and model accuracy [44]. Similarly, Snavey *et al.* [20] constructed a skeletal graph to process a large number of unordered collections. This process dramatically improved efficiency. Another suggestion is to simplify the network while employing entire images in reconstruction. We introduced the image topology skeleton technique [25] by selecting the image topology number as the weight to reduce network complexity. Nevertheless, studies on the applicability of the integration of the given methods are required for large-scale scene geometry reconstruction.

5. Conclusions

This study presents a new methodology for reconstructing 3D models. The new method is applied to historical heritage objects. This method combines terrestrial laser scanning and image reconstruction techniques. We mainly focus on the 3D reconstruction of huge or complex monuments which possesses constraints for data acquisition with any single sensor. Moreover, we exploit integration at the data level instead of at the model level to overcome the weakness of each data source [1]. The proposed methodology is tested on a historical monument with complex shapes and results in a maximum of 12‰ model error of relative positioning.

The open-source Bundle is selected as the basic SfM to process the UAV optical collections. In the image matching procedure, we propose a camera network while considering camera position and attitude to simplify image tracks. In the analysis, we consider a camera network with 45 images captured with the use of an eight-rotor UAV platform carrying a Canon 5D Mark II camera. The performance is tested, and a reduced number of 564 image tracks were obtained when compared with the 990 traversal tracks (a nearly 50% reduction). Figure 5 highlights the positive performance of our camera network matching method, which successfully reduced the number of mismatches that emerged in traversal matching. Further, the comparison of the 3D point clouds in Table 1 reveals that the camera network is satisfactory in terms of efficiency and completeness. Figure 5 shows the final 3D points produced from the camera network modified SfM. Although, accuracy is not the primary goal, the camera network achieved acceptable model accuracy with 12‰ model error.

The TLS cloud is characterized by high sensing density in our study. Thus, such cloud can be used to enhance details. In addition, a UAV image-reconstructed cloud covers the overall information of scene geometry. The limited accuracy of the image-level cloud is considered sufficient to reveal surface shape and structure. The integration of the two groups of point clouds improves the coverage completeness, which enables the modeling of the details and the entirety of complex heritage objects. Perspective images captured through airborne and terrestrial sensing can be used to obtain a photorealistic 3D model. However, this topic is outside the main focus of this paper.

Further work can investigate the integration of multi-scale airborne sensing to obtain a more reliable 3D model in terms of the scene resolution of image-level point clouds. Ultimately, future work should consider the improvement of multi-source data fusion to solve issues resulting from resolution differences.

Author Contributions

Zhihua Xu designed the camera network modified structure from motion pipeline, integrated the two groups of point clouds from both image-reconstruction and terrestrial laser scanning, interpreted the results and wrote the manuscript. Lixin Wu conceived the study, designed and guided the experiment, and contributed to manuscript writing and revision. Yonglin Shen, Fashuai Li and Ran Wang performed the field experiments. Yonglin Shen advised on the design of the camera network construction. Fashuai Li and Qiuling Wang draw the final 3D models of the monument with the integrated point cloud. Ran Wang developed the 3D reconstruction software according to both the camera network and open source code Bundler.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Remondino, F. Heritage recording and 3D modeling with photogrammetry and 3D scanning. *Remote Sens.* **2011**, *3*, 1104–1138.
2. Yastikli, N. Documentation of cultural heritage using digital photogrammetry and laser scanning. *J. Cult. Herit.* **2007**, *8*, 423–427.
3. EI-Hakim, S.F.; Beraldin, J.A.; Picard, M.; Godin, G. Detailed 3D reconstruction of large-scale heritage sites with integrated techniques. *IEEE Comput. Graph. Appli.* **2004**, *24*, 21–29.
4. Remondino, F.; Rizzi, A.; Agugiaro, G.; Girardi, S.; de Amicis, R.; Magliocchetti, D.; Girardi, G.; Baratti, G. Geomatics and geoinformatics for digital 3D documentation, fruition and valorization of cultural heritage. In Proceedings of the EUROMED 2010 Workshop "Museum Futures: Emerging Technological and Social Paradigms", Lemessos, Cyprus, 8–13 November 2010.
5. CARARE Best Practice Network. Available online: http://www.carare.eu (accessed on 29 June 2014).
6. 3D-ICONS Pilot Project. Available online: http://3dicons-project.eu (accessed on 29 June 2014).
7. 3D-COFORM Consortium. Available online: http://3dcoform.eu (accessed on 29 June 2014).
8. Guarnieri, A.; Remondino, F.; Vettore, A. Digital photogrammetry and TLS data fusion applied to cultural heritage 3D modeling. In Proceedings of the ISPRS Commission V Symposium "Image Engineering and Vision Metrology", Dresden, Germany, 25–27 September 2006; Maas, H.-G., Schneider, D., Eds.; Volume XXXVI, Part 5.
9. Böhler, W.; Marbs, A. 3D scanning instruments. In Proceedings of the CIPA WG 6 International Workshop on Scanning for Cultural Heritage Recording, Corfu, Greece, 1–2 September 2002.
10. Pieraccini, M.; Guidi, G.; Atzeni, C. 3D digitizing of cultural heritage. *J. Cult. Herit.* **2001**, *2*, 63–70.
11. Koutsoudis, A.; Vidmar, B.; Ioannakis, G.; Arnaoutoglou, F.; Pavlidis, G.; Chamzas, C. Multi-image 3D reconstruction data evaluation. *J. Cult. Herit.* **2013**, *15*, 73–79.

12. Gruen, A.; Huang, T.S. *Calibration and Orientation of Cameras in Computer Vision*; Springer Berlin-Heidelberg: Berlin, Germany, 2001; p. 239.

13. Remondino, F.; El-Hakim, S. Image-based 3D modelling: A review. *Photogramm. Rec.* **2006**, *21*, 269–291.

14. Robertson, D.; Cipolla, R. Structure from motion. In *Practical Image Processing and Computer Vision*; John Wiley: Hoboken, NJ, USA, 2009.

15. Furukawa, Y.; Ponce, J. Accurate, dense, and robust multiview stereopsis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 1362–1376.

16. Engels, C.; Stewénius, H.; Nistér, D. Bundle adjustment rules. In Proceedings of Photogrammetric Computer Vision (PCV), Bonn, Germany, 20–22 September 2006.

17. Snavely, N. Bundler: Structure from Motion for Unordered Image Collections. 2010. Available online: http://www.cs.cornell.edu/~snavely/bundler (accessed on 29 June 2014).

18. Snavely, N.; Seitz, S.M.; Szeliski, R. Photo tourism: Exploring photo collections in 3D. *ACM Trans. Graph.* **2006**, *25*, 835–846.

19. Wu, C.; Agarwal, S.; Curless, B.; Seitz, S.M. Multicore bundle adjustment. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence, RI, USA, 20–25 June 2011; pp. 3057–3064.

20. Snavely, N.; Seitz, S.M.; Szeliski, R. Skeletal graphs for efficient structure from motion. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Anchorage, AK, USA, 23–28 June 2008.

21. Wu, C. VisualSFM: A Visual Structure from Motion System. 2011. Available online: http://ccwu.me/vsfm (accessed on 29 June 2014).

22. PhotoModeler Scanner. Available online: http://www.photomodeler.Com/products/pm-scanner.htm (accessed on 29 June 2014).

23. Agisoft PhotoScan. Available online: http://www.agisoft.ru (accessed on 29 June 2014).

24. Pix4D Simply Powerful. Available online: http://pix4d.com (accessed on 29 June 2014).

25. Xu, Z.; Wu, L.; Wang, Z.; Wang, R.; Li, Z.; Li, F. Matching UAV images with image topology skeleton. In Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS), Melbourne, VIC, Australia, 21–26 July 2013.

26. Xu, Z.; Wu, L.; Chen, S.; Wang, R.; Li, F.; Wang, Q. Extraction of image topological graph for recovering the scene geometry from UAV collections. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2014**, *XL*, 319–323.

27. Alsadik, B.; Gerke, M.; Vosselman, G. Automated camera network design for 3D modeling of cultural heritage objects. *J. Cult. Herit.* **2013**, *14*, 515–526.

28. Remondino, F.; Rizzi, A. Reality-based 3D documentation of natural and cultural heritage sites—Techniques, problems, and examples. *Appl. Geomat.* **2010**, *2*, 85–100.

29. Lingua, A.M.; Rinaudo, F. The statue of Ramsete II: Integration of digital photogrammetry and laser scanning technique for 3D modelling. In Proceedings of the 18th International Symposium CIPA, Potsdam, Germany, 18–21 September 2001; pp. 206–211.

30. Ioannidis, C.; Tsakiri, M. Laser scanning and photogrammetry for the documentation of a large statue-experiences in the combined use. In Proceedings of the CIPA XIX International Symposium, Athens, Greece, 1–6 October 2003.

31. Beraldin, J.-A. Integration of laser scanning and close-range photogrammetry-the last decade and beyond. In Proceedings of the XXth ISPRS Congress, Commission VII, Istanbul, Turkey, 12–23 July 2004; pp. 972–983.

32. Böhler, W.; Marbs, A. 3D Scanning and photogrammetry for heritage recording: A comparison. In Proceedings of the 12th International Conference on Geoinformatics, Gävle, Sweden, 7–9 June 2004; pp. 291–298.

33. Kadobayashi, R.; Kochi, N.; Otani, H.; Furukawa, R. Comparison and evaluation of laser scanning and photogrammetry and their combined use for digital recording of cultural heritage. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2004**, *XXXV*, 401–406.

34. Remondino, F.; Guarnieri, A.; Vettore, A. 3D modeling of close-range objects: Photogrammetry or laser scanning? *Proc. SPIE* **2005**, doi:10.1117/12.586294.

35. Lerma, J.L.; Navarro, S.; Cabrelles, M.; Seguí, A.E.; Haddad, N.; Akasheh, T. Integration of laser scanning and imagery for photorealistic 3D architectural documentation. In *Laser Scanning, Theory and Applications*; Wang, C.C., Ed.; InTech: Rijeka, Croatia, 2011; pp. 413–430.

36. Pollefeys, M.; van Gool, L.; Vergauwen, M.; Verbiest, F.; Cornelis, K.; Tops, J.; Koch, R. Visual modeling with a hand-held camera. *Int. J. Comput. Vis.* **2004**, *59*, 207–232.

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

38. Douterloigne, K.; Gautama, S.; Philips, W. On the accuracy of 3D landscapes from UAV image data. In Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS), Honolulu, HI, USA, 25–30 July 2010.

39. Besl, P.J.; McKay N.D. Method for registration of 3-D shapes. *IEEE Trans. Pattern Anal. Mach. Intell.* **1992**, *14*, 239–256.

40. Pulli, K. Multiview registration for large data sets. In Proceedings of the 2th International Conference on 3-D Digital Imaging and Modeling (3DIM'99), Ottawa, ON, Canada, 4–8 October 1999; pp. 160–168.

41. Bae, K.-H.; Lichti, D.D. A method for automated registration of unorganized point clouds. *ISPRS J. Photogramm. Remote Sens.* **2008**, *63*, 36–54.

42. González-Aguilera, D.; Rodríguez-Gonzálvez, P.; Gómez-Lahoz, J. An automatic procedure for co-registration of terrestrial laser scanners and digital cameras. *ISPRS J. Photogramm. Remote Sens.* **2009**, *64*, 308–316.

43. Weinmann, M.; Hinz, S.; Jutzi, B. Fast and automatic image-based registration of TLS data. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 62–70.

44. Alsadik, B.; Gerke, M.; Vosselman, G.; Daham, A.; Jasim, L. Minimal camera networks for 3D image based modeling of cultural heritage objects. *Sensors* **2014**, *14*, 5785–5804.