Multi-source data-based 3D digital preservation of large-scale ancient Chinese architecture: A case report

Xiang GAO1,2, Hainan CUI1, Lingjie ZHU1,2, Tianxin SHI1,2, Shuhan SHEN1,2

1. National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
2. University of Chinese Academy of Sciences, Beijing 100049, China

* Corresponding author, shshen@nlpr.ia.ac.cn
Received: 7 May 2019    Accepted: 2 August 2019

Supported by the National Natural Science Foundation of China (61632003, 61873265).

Abstract   The 3D digitalization and documentation of ancient Chinese architecture is challenging because of architectural complexity and structural delicacy. To generate complete and detailed models of this architecture, it is better to acquire, process, and fuse multi-source data instead of single-source data. In this paper, we describe our work on 3D digital preservation of ancient Chinese architecture based on multi-source data. We first briefly introduce two surveyed ancient Chinese temples, Foguang Temple and Nanchan Temple. Then, we report the data acquisition equipment we used and the multi-source data we acquired. Finally, we provide an overview of several applications we conducted based on the acquired data, including ground and aerial image fusion, image and LiDAR (light detection and ranging) data fusion, and architectural scene surface reconstruction and semantic modeling. We believe that it is necessary to involve multi-source data for the 3D digital preservation of ancient Chinese architecture, and that the work in this paper will serve as a heuristic guideline for the related research communities.

Keywords    Ancient Chinese architecture; 3D digital preservation; Multi-source data acquisition; Architectural scene modeling

1 Introduction

Along with European and Islamic architectures, ancient Chinese architecture is one of the most important components of the world architectural system, with its most significant characteristic being the use of a timber framework. Though it allows for more delicate structures to be achieved, as compared with other architectural styles, ancient Chinese architecture is more vulnerable to natural disasters, e.g., fire or earthquake. As a result, there is an urgent need to preserve ancient Chinese architecture and one of the best means to achieve this is to digitally preserve it by reconstructing its complete and detailed 3D models.

Architectural scene modeling has always been an intensive research topic in the fields of computer vision, computer graphics, and photogrammetry. Though many exciting studies have been conducted, most of them have performed the modeling task using single-source data. Some methods generate the scene models from images with similar viewpoint and scale only[1-4], which are captured either using handheld cameras or cameras mounted on unmanned aerial vehicles (UAVs), while others try to obtain the models by using range data, e.g., RGB-D images[5-8] obtained from Kinect or LiDAR data[9-12] acquired using a laser
However, it is difficult to generate accurate and complete architectural scene models by using single-source data, especially for ancient Chinese architecture with complicated structures. In this study, we use multisource data acquisition, processing, and fusion to achieve 3D digital preservation of ancient Chinese architecture. There are four types of acquired data: (1) aerial images captured by an interchangeable lens digital (ILD) camera mounted on a UAV; (2) ground images captured by a digital single lens reflex (DSLR) camera mounted on a robotic camera mount; (3) laser point cloud scanned by a laser scanner; and (4) geo-coordinates of the ground control points (GCPs) measured by a differential GPS system. The first three data types are used for scene modeling and complementarity: the aerial and ground images provide large-scale and close-range capturing of a scene, while the images and LiDAR data are complementary in flexibility and accuracy. The third data type is used for image geo-reference and accuracy evaluation.

In the following sections, we first briefly introduce the ancient Chinese temples we surveyed. Then, we report the used data acquisition equipment and the acquired multisource data. Finally, we give an overview of the several applications we conducted based on the acquired data.

2 Scenes

The architectural scenes we surveyed in this paper are two ancient Chinese temples, named Foguang Temple (FGT) and Nanchan Temple (NCT) (Figure 1), which are two of the four existing Chinese architectures with timber structure built in Tang Dynasty. Among them, FGT is the largest one, while NCT is the oldest one. These two temples are described below.

Figure 1  Multi-view stereo (MVS) point clouds of FGT and NCT, which are generated using the method proposed in [13]. The red, green, and blue rectangles in the left figure denote the Great East Hall (GEH), Manjusri Hall (MJH), and Garan Hall (GRH) of FGT. The red rectangle in the right figure denotes the Great Buddha Hall (GBH) of NCT.

2.1 Foguang Temple¹

FGT is a Buddhist temple located in Wutai County, Shanxi Province of China, covering an area of approximately 34000m². It mainly contains three halls: The Great East Hall (GEH), Manjusri Hall (MJH), and Garan Hall (GRH).

2.1.1 Great East Hall

Dating from 857 of the Tang Dynasty, the GEH is the third oldest but the largest wooden building in China. The hall is located on the far-east side of the temple. It is a single-story structure, measuring seven bays by four, and is supported by inner and outer sets of columns. On top of each column is a complicated set of brackets containing seven different bracket types. Inside the hall are 36 sculptures, as well as murals on

¹https://en.wikipedia.org/wiki/Foguang_Temple
each wall, which date from the Tang Dynasty and later periods.

2.1.2 Manjusri Hall
On the north side of the temple courtyard is the MJH, which was constructed in 1137 during the Jin Dynasty and is roughly the same size as the GEH. It features a single-eave hip gable roof. The interior of the hall has only four support pillars. To support the large roof, diagonal beams are used. On each of the four walls are murals of arhats painted in 1429 during the Ming Dynasty.

2.1.3 Garan Hall
The GRH is located in the southwest corner of FGT. It was first built in 1628 of the Ming Dynasty and rebuilt in 1661 of the Qing Dynasty, measuring three bays in width. The sculptures of 18 Gran Gods are located in the GRH.

2.2 Nanchan Temple
NCT is a Buddhist temple, also located in Wutai County, Shanxi Province of China, covering an area of approximately 3100m². It contains only one main hall, named Great Buddha Hall (GBH).

2.2.1 Great Buddha Hall
Built in 782 during the Tang Dynasty, the GBH is the oldest preserved timber building extant of China. It is a three-bay square hall. Not only is GBH an important architectural structure, but it also contains an original set of artistically important Tang sculptures dating from the period of its construction. Seventeen sculptures share the interior space of the hall with a small stone pagoda.

3 Equipment
The equipment for data acquisition are divided into four categories according to the data types, i.e., equipment for (1) aerial images; (2) ground images; (3) LiDAR data acquisition; and (4) GCPs (Figure 2). In the following subsections, we introduce and briefly describe the equipment we used.

3.1 Equipment for aerial image acquisition
For aerial image acquisition, we used an ILD camera, Sony NEX-5R, mounted on a UAV, Microdrones Md4-1000.

3.1.1 Sony NEX-5R
The Sony NEX-5R is an ILD camera with a 16.1 effective megapixel complementary metal oxide semiconductor sensor. It has a similar imaging quality, but much lighter weight as compared with standard DSLR camera. The above features make the Sony NEX-5R more suitable for aerial image capturing.

3.1.2 Microdrones Md4-1000
The Microdrones Md4-1000 system is a leading vertical take-off and landing, autonomous unmanned micro aerial vehicle. The drone body and camera mount are made of carbon fiber material, which is lighter in weight and higher in strength.

---

3 https://www.sony.com/  
4 https://www.microdrones.com/
3.2 Equipment for ground image acquisition

For ground image acquisition, we used a DSLR camera, Canon EOS 5D Mark III, mounted on a robotic camera mount, GigaPan Epic Pro.

3.2.1 Canon EOS 5D Mark III*

The Canon EOS 5D Mark III is one of the most famous DSLR cameras. It is equipped with a 22.3 megapixel full-frame CMOS sensor and has excellent imaging quality under various environments. As a result, it is a suitable choice for scene capturing from the ground.

3.2.2 GigaPan Epic Pro*

The GigaPan Epic Pro is a robotic camera mount that can capture HD, gigapixel photos using almost any digital camera. By setting the upper left and lower right corners of the desired panorama, the camera mount determines the number of photos the camera needs to take, and then, automatically organizes them.

3.3 Equipment for LiDAR data acquisition

For LiDAR data acquisition, we used a laser scanner, Leica ScanStation P307. The Leica ScanStation P30 delivers high-quality 3D data and high-dynamic-range imaging at an extremely fast scan rate of 1 million points per second and within a range of up to 270m with extremely high accuracy. For example, its 3D position accuracy is 3mm at 50m and 6mm at 100m.

3.4 Equipment for GCP measurement

For GCP measurement, we used a differential GPS system, Hi-Target V30 GNSS RTK8. The V30 GNSS

---

* https://www.usa.canon.com/
* http://www.gigapan.com/
* http://www.leica-geosystems.com/
RTK possesses outstanding positioning performance. For example, its horizontal and vertical positioning accuracies in a high-precision static situation are 2.5mm+0.1ppm root mean square (RMS) and 3.5mm+0.4ppm RMS, respectively.

4 Data

In this section, we introduce the multisource data acquired in the scenes described in Sec. 2 with the equipment described in Sec. 3. The acquired multisource data comprise aerial images, ground images, LiDAR data, and GCPs.

4.1 Aerial images

We manually flew Microdrones Md4-1000 over FGT and NCT and triggered the Sony NEX-5R shutter to capture aerial images. The images were captured with five flight paths, one for nadir images and the other four for 45° oblique images. These images had a resolution of 4912×3264 pixels. We took 1596 and 772 aerial images of FGT and NCT, respectively. Examples of the aerial images are shown in Figure 3. In addition, the structure from motion (SfM) point clouds and camera poses of the FGT and NCT aerial images were computed using the method proposed in [1], and are shown in Figure 4.

![Aerial image examples of FGT and NCT](image)

Figure 3 Aerial image examples of FGT and NCT.

![Aerial SfM point cloud of FGT and NCT](image)

Figure 4 Aerial SfM point clouds of FGT and NCT generated using the method proposed in [1]. The red cones denote the poses of aerial cameras.

4.2 Ground images

We mounted the Canon EOS 5D Mark III on the Gigapan Epic Pro and took ground images station-by-station. The Gigapan Epic Pro was set to capture images with a pitch range of −40°–40°, step 20°, and yaw
range of $0^\circ - 320^\circ$, step $40^\circ$, which allowed for 45 ground images to be captured for each station. The captured ground images had a resolution of $5760 \times 3840$. There were 155, 55, 32, and 6 image-capturing stations for the outdoor scenes of FGT and indoor scenes of the GEH, MJH, and GRH, respectively. In addition, there were 62 and 19 image-capturing stations for the outdoor scenes of the NCT and indoor scenes of the GBH, respectively. The ground image examples of FGT and NCT are shown in Figures 5 and 6, respectively. In addition, the SfM point clouds and camera poses of the FGT and NCT ground images, shown in Figure 7, were computed using the method proposed in [1].

**4.3 LiDAR data**

We used the Leica ScanStation P30 to acquire LiDAR data of FGT and NCT. Prior to scanning, we determined the laser scanning locations. There were 39, 35, 16, and 3 laser scanning stations for outdoor scenes of FGT and indoor scenes of the GEH, MJH, and GRH, respectively. In addition, there were 12 and 8 laser scanning stations for outdoor scenes of the NCT and indoor scenes of the GBH, respectively. For each station, we obtained approximately 100 million high-accuracy laser points with RGB information. The laser point cloud examples of FGT and NCT are shown in Figure 8, and the locations of the outdoor laser scanning stations of FGT and NCT are shown in Figure 9.
Ground control points

The geo-coordinates of the GCPs were measured using the V30 GNSS RTK system. The GCPs had two usages in this study: (1) for geo-referring the (aerial and ground) images; and (2) serving as ground truths for evaluating the calibration results of the (aerial and ground) cameras. There are two types of GCPs according to the camera type: (1) GCPs for aerial cameras, which are manually selected in the scenes and marked in the aerial images, and thus, are usually obvious corners. There are 53 and 33 GCPs of this type for FGT and NCT, respectively, and Figures 9 and 10 provide several examples of these. (2) GCPs for
outdoor ground cameras, which are located at the outdoor image-capturing stations to accurately record their geo-coordinates. As a result, there are 155 and 62 GCPs of this type for FGT and NCT, respectively, which are the same as the number of outdoor image-capturing stations.

5 Applications

Based on the acquired multisource data, we mainly conducted four types of applications: (1) aerial and ground image fusion\cite{13}; (2) image and LiDAR data fusion\cite{14}; (3) surface reconstruction\cite{15}; and (4) semantic modeling\cite{16}. These are introduced below.

5.1 Aerial and ground image fusion

To reconstruct a complete 3D digital model of ancient Chinese architecture that captures details of complex structures, e.g., cornices and brackets, usually two image sources, aerial and ground, are involved in large-scale and close-range scene capturing. When using both aerial and ground images, a common practice is to conduct the reconstruction separately to first generate aerial and ground point clouds and then fuse them. Considering the noisy nature of the reconstructed 3D point clouds from image collections and the loss of rich textural and contextual information of 2D images in 3D point clouds, it is preferable to fuse the point clouds via 2D image feature point matching rather than by direct 3D point cloud registration, e.g., iterative closest point (ICP)\cite{17}. To fuse the aerial and ground images for complete scene model reconstruction, two issues should be specially addressed: (1) how to match the aerial and ground images with substantial variations in viewpoint and scale; and (2) how to fuse the aerial and ground point clouds with drift phenomena and notable differences in noise level, density, and accuracy.
To deal with the aerial and ground image matching problem, in [13], the ground image was warped to the viewpoint of the aerial image, through which the differences in the viewpoint and scale between these two types of images were eliminated. Unlike the method proposed in [18], which synthesizes the aerial-view image by leveraging the spatially discrete ground MVS point cloud, the image synthesis method proposed in [13] resorts to the spatially continuous ground sparse mesh, which is reconstructed from the ground SfM point cloud. For a pair of aerial and ground images, each spatial facet in their co-visible ground sparse mesh induced a local homography between them. The aerial-view image was synthesized by warping the ground image to the aerial one using the induced homographies. Note that the above image synthesis method is free from the time-consuming MVS procedure and the resultant synthesized images do not suffer from missing pixels in the co-visible regions of aerial and ground image pairs.

After image synthesis, the synthesized image was matched with the target aerial image by using the scale-invariant feature transform (SIFT) feature point extraction and matching. In [13], instead of filtering out the inevitable point match outliers by the nearest-neighbor distance ratio test [20] which is prone to discarding true positives, this was achieved using the following two techniques: (1) a consistency check of the feature scales and principal orientations between the point matches; and (2) an affine transformation verification of the feature locations between the point matches. Note that, unlike the commonly used fundamental matrix-based outlier filtering scheme, which provides a point-to-line constraint, the affinity-based scheme in [13] provides a point-to-point constraint, and thus, is more effective for outlier filtering. Figure 11 gives an image feature-matching example of a pair of aerial and ground images.

To tackle the aerial and ground point cloud fusion issue, rather than align the point clouds by estimating a similarity transformation [21] between them with random sample consensus [22], which was performed in [18, 23, 24], the point clouds were fused by a global bundle adjustment (BA) [25] to deal with the possible scene drift phenomenon. To achieve this, in [13], the obtained aerial and ground point matches were first linked to the original aerial tracks. Figure 12 shows a cross-view track linking example. Then, a global BA was performed to fuse the aerial and ground SfM point clouds with the augmented aerial tracks and the original ground tracks. Figure 13 shows the aerial and ground SfM point cloud fusion results of FGT and NCT.

5.2 Image and LiDAR data fusion

There are two key issues in reconstructing large-scale architectural scenes: accuracy and completeness. Though many existing methods focus on the issue of reconstruction accuracy, they pay less attention to reconstruction completeness. When the architectural scene is complicated, e.g., ancient Chinese architecture,
the reconstruction completeness of the common pipelines is difficult to guarantee. To reconstruct accurate and complete 3D models (point clouds or surface meshes) of the large-scale and complicated architectural scenes, both global structures and local details of the scenes need to be surveyed. Currently, there are two frequently used surveying methods for scene reconstruction: image-based\[1-4\] and laser scanning-based\[9-12\]. These two approaches are complementary in flexibility and accuracy: the image-based reconstruction methods are convenient and flexible, but heavily depend on several external factors, e.g., illumination variation, textural richness, and structural complexity, while the laser scanning-based reconstruction methods possess high accuracy and are robust to adverse conditions, but are expensive and time-consuming.

To generate a complete scene reconstruction by fusing images and LiDAR data, a straightforward approach is to treat images and LiDAR data equally. Specifically, architectural scene models are obtained from these two types of data, respectively, at first and fused with GCPs\[26\] or using ICP algorithm\[27,28\] afterward. However, this is nontrivial because the point clouds generated from images and laser scans have significant differences in density, accuracy, completeness, etc., which result in inevitable registration errors. In addition, the laser scanning locations need to be carefully selected to guarantee the scanning
overlap for their self-registration.

To address the above issues, we proposed a more effective data collection and scene reconstruction pipeline in [14], which considers both data collection efficiency and reconstruction accuracy and completeness. Our pipeline uses images as primacy to completely cover the scene, and laser scans as a supplement to deal with low-textured, low-lighting, or complicated structured regions. Similar to [13], in [14], images and LiDAR data were fused by 2D image feature point matching between the captured images and images synthesized from LiDAR data, instead of 3D point cloud registration.

In [14], we first obtained a fused SfM point cloud from the captured aerial and (outdoor and indoor) ground images. For this purpose, both point matches between aerial and ground images and between outdoor and indoor images are required. However, obtaining these two types of point matches is non-trivial, due to (1) the large viewpoint and scale differences between the aerial and ground images and (2) the limited view overlapping between the outdoor and indoor images. In [14], we first generated SfM point clouds from the aerial, outdoor, and indoor images individually, and then, fused them using cross-view point matches. The aerial and ground point matches were obtained using the method proposed in [13], while the outdoor and indoor point matches were obtained by matching the outdoor and indoor images near the door.

Then, the aerial- and ground-view synthesized images were generated from the laser point clouds and matched with the captured ones to obtain cross-domain correspondences. Figures 14 and 15 show image feature-matching examples of a pair of synthesized aerial-view and captured aerial images and a pair of synthesized ground-view and captured ground images, respectively. Based on the cross-domain 2D point matches, the images and LiDAR data were fused in a coarse-to-fine scheme. The laser point cloud of each scanning station was coarsely registered to the fused SfM point cloud individually by a similarity transformation[21] between them, which was estimated using RANSAC[22]. The 3D point correspondences for similarity transformation estimation were converted from the obtained cross-domain 2D point matches. Then, the camera poses of the captured images, the spatial coordinates of the SfM point cloud, and the alignments of the laser scans were jointly optimized by a generalized BA to finely merge the images.
and LiDAR data. The BA procedure shown in [14] is called a generalized one because the camera poses and laser scan alignments were simultaneously optimized by minimizing both 2D–3D reprojection errors and 3D–3D space errors. Figure 16 shows the SfM and laser point cloud fusion results of FGT and NCT.

![Figure 16](image)

**Figure 16** SfM and laser point cloud fusion results of FGT and NCT by using the method proposed in [26]. The first row: result of NCT, the second row: result of FGT. From left to right: (aerial-outdoor-indoor) SfM point clouds, (outdoor-indoor) laser point clouds, and fused SfM and laser point clouds (red for laser point clouds, green for aerial SfM point clouds, and blue for ground SfM point clouds).

### 5.3 Surface reconstruction

Though tremendous progress has been made in the community of image-based architectural scene reconstruction recently, in the case of large-scale scenes with multi-scale objects, the current reconstruction methods have some problems in terms of completeness and accuracy, especially when concerning scene details.

Scene details such as small-scale objects and object edges are essential parts of scene surfaces. Figure 17 shows an example of preserving the scene details in reconstructing FGT. In general, representing scene details, e.g., the brackets in Figure 17, in cultural heritage digitalization projects is among the most important tasks. Point cloud representation is often redundant and noisy, while mesh representation is concise, but it sometimes loses some information. Therefore, preserving scene details in reconstructing multi-scale scenes has been a difficult problem in surface reconstruction. The existing surface reconstruction methods[29–32] either ignore the scene details or rely on further refinement to restore them. This is attributable to the following reasons. First, compared to noise, the supportive points in such parts of the scene are sparse, making it difficult to distinguish true surface points from false ones. Second, the visibility models and associated parameters employed in the existing methods are not particularly suitable for large-scale ranges, where scene details are usually compromised in terms of overall accuracy and completeness. As the first case seems to be unsolvable because of a lack of sufficient information, we focus on the second one in [15].

![Figure 17](image)

**Figure 17** Surface reconstruction result (right) of FGT from redundant and noisy MVS point cloud (left) by using the method proposed in [15].
to generate accurate surface meshes. To use the visibility information, assumptions of the visibility model are made so that the space between the camera center and the 3D point is free, while that behind the point along the line of sight is full. However, the above visibility model has two shortcomings: (1) the points are often contaminated with noise; and (2) the full-space scales are often difficult to determine. To address these issues, the main work and contributions of our method in [15] are three-fold. (1) To preserve scene details without decreasing the noise filtering ability, we propose a new visibility model with error tolerance and adaptive end weights. (2) We also introduce a new likelihood energy representing the punishment of wrongly classifying a part of space as free or full, which helps to improve the ability of the proposed method to efficiently filter noise (Figure 18). (3) We further improve the performance of the proposed method by using the dense visibility technique, which helps to keep the object edge sharp (Figure 19).

![Figure 18](image1.png) Surface reconstruction without and with the likelihood energy proposed in [15]. From left to right: MVS point cloud with heavy noise, reconstructed meshes without and with the likelihood energy.

![Figure 19](image2.png) Surface reconstruction without and with the dense visibility technique proposed in [27]. From left to right: original image and its depth map, reconstructed meshes without and with the dense visibility technique.

### 5.4 Semantic modeling

3D semantic modeling from images has gained popularity in recent years. Its goal is to obtain both the 3D structure and semantic knowledge of a scene. 3D semantic models help humans and automatic systems know "what" is "where" in a specific scene, which is a stated goal of computer vision and has various applications in fields such as automatic piloting, augmented reality, and service robotics. Over the last decade, tremendous progress has been made in the field of 3D geometric reconstruction, which enables us to reconstruct large-scale scenes with high level of detailing. At the same time, deep learning techniques have led to a huge boost in 2D image understanding, such as semantic segmentation and instance recognition. Thus, the combination of deep learning and geometry reconstruction to acquire 3D semantic models is attracting increasing research interest nowadays. Generally, there are two methods of achieving this goal: (1) jointly optimizing the 3D structure and semantic meaning of the scene[31,33,34], and (2) assigning semantic labels to the estimated 3D structure[35-37]. Our work in [16] falls in the second category, i.e., we focus on labeling the existing 3D geometry models, especially fine-level labeling of large-scale mesh...
Using the state-of-the-art SfM\(^1,3^{8}\) and MVS\(^2,3^{2}\) algorithms, a detailed 3D model could be reconstructed from hundreds and thousands of images. A straightforward way to label this model is to annotate each facet directly. However, this process is quite cumbersome because there is no effective tool for manual annotation in 3D space, and the current deep learning-based labeling pipeline, such as that proposed in \([39, 40]\), cannot deal with large-scale 3D models. Thus, a feasible method for large-scale 3D model labeling is to first perform pixel-wise semantic segmentation on 2D images, and then back-project these labels into 3D space using the calibrated camera parameters and fuse them. Apparently, in this way, the quality of the 3D semantic labeling highly depends on that of the 2D semantic segmentation. Current 2D semantic segmentation methods tend to fine-tune a pre-trained convolutional neural network (CNN) within the transfer learning framework, but still require several manually annotated images for cross-domain datasets. However, in specialized domains, such as fine-level labeling of ancient Chinese architecture, only experts with special knowledge and skills can annotate them reliably. Therefore, reducing the cost of annotation is meaningful. In \([16]\), we proposed a novel method that can dramatically reduce the annotation cost by integrating active learning (AL) into the fine-tuning process. AL is an established way to reduce the labeling workload by iteratively choosing images for annotation to train the classifier for better performance.

In \([16]\), we started by fine-tuning a CNN for image semantic segmentation with a limited number of annotated images, and used it to segment all other unannotated images. Then, all predicted image labels were back-projected into 3D space and fused on the 3D model using the Markov random field. As the 3D semantic model considers both 2D image segmentation and 3D geometry, it can be used as a reliable intermediate to select the most worthy image candidates for annotation, and then proceed the next fine-tune iteration. This training-fusion-selection process continues until the label configuration of the model becomes steady. Figure 20 shows the pipeline of our method proposed in \([16]\), and Figure 21 shows the semantic modeling results of FGT and NCT.

![Figure 20 Pipeline of the method proposed in [37].](image)

6 Conclusions

This paper reports our work on the 3D digital preservation of large-scale ancient Chinese architecture based on multi-source data. We first introduce two famous ancient Chinese temples we surveyed, FGT and NCT. Then, we briefly introduce the data acquisition equipment we used: (1) Sony NEX-5R and Microdrones Md4-1000 for aerial images; (2) Canon EOS 5D Mark III and GigaPan Epic Pro for ground images; (3) Leica ScanStation P30 for LiDAR data; and (4) Hi-Target V30 GNSS RTK for GCPs. Subsequently, we report the multi-source data acquired using the above equipment and show several
examples of these. Finally, we provide an overview of several applications of the multi-source data, including ground and aerial image fusion\cite{13}, image and LiDAR data fusion\cite{15}, and architectural scene surface reconstruction\cite{27} and semantic modeling\cite{16}. We believe that involving multi-source data is a more effective way for the 3D digital preservation of ancient Chinese architecture, and that this paper can serve as a heuristic guideline for related research communities.

References


Figure 21 Semantic modeling results of FGT and NCT by using the method proposed in [37]. The first row: result of NCT, the second row: result of FGT.
Proceedings of the
Szeliski R
Seitz S M
Agarwal S
Snavely N
Bundle adjustment in the large
In
th European
Gao X
Shen S H
Pattern
Hu Z Y
,.
,.
,.
,.
IEEE Transactions on
Least-squares estimation of transformation parameters between two point patterns
IEEE Transactions on Pattern Analysis and
Schmid C
A performance evaluation of local descriptors
Mikolajczyk K
Lowe D G


