Large Scale and Complex Structure Grotto Digitalization Using Photogrammetric Method: A Case Study of Cave No. 13 in Yungang Grottoes

Hanyu Xiang[1, 2*], Wenyuan Niu[1, 2*], Xianfeng Huang[1, 2, 4, 5†], Bo Ning[3], Fan Zhang[1, 2, 4, 5], Jianmin Xu[5]

[1]State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China
[2]Intelligent Computing Laboratory for Cultural Heritage, Wuhan University, Wuhan, China
[3]Yungang Research Institute, Datong, China
[4]Archaeological Institute for Yangtze Civilization, Wuhan University, Wuhan, China
[5]Daspatial Technology Co., Ltd., Wuhan, China

Keywords: 3D Modeling, Cultural Heritage, Digitalization, Large Scale, Multi-view Images, Grottoes.

Abstract

3D reconstruction of cultural heritage with large volume and high precision is a technical problem in the field of photogrammetry. This paper studies a high-precision digitalization method for large-volume immovable heritage assets based on photogrammetry and laser scanning. It solves the problem of large-scale aerial triangulation and ensures overall color and geometric consistency while satisfying high-precision modeling of local details. Taking the millimeter accuracy 3D reconstruction project of Cave No. 13 in Yungang Grottoes as an example, we use more than 280,000 arbitrary images to reconstruct the entire cave and verify the effectiveness of the proposed method.

1. Introduction

Digitalization of heritage assets is an important research direction in the field of photogrammetry, through which the preservation and display of heritage assets can be realized. As a typical immovable cultural heritage asset, grottoes contain richly painted murals and complex geometric structures. Therefore, it is often necessary to use multiple sensors to achieve high-precision 3D reconstruction.

Photogrammetry is an important method to document and preserve heritage objects and sites (Jebur, 2022). Within this framework, close-range photogrammetry and aerial photogrammetry by unmanned vehicles (UAVs) enable 3D reconstruction in different levels of detail by exploiting information obtained from images (De Fino et al., 2023). There are two main aspects in the digital application of photogrammetry technology in cultural heritage: the digitalization of heritage assets in collections and immovable heritage assets. In digitalizing heritage assets in collections, a relatively controllable collection environment is constructed, and close-range photogrammetry is used to complete 3D reconstruction. Nicolae et al. showed how close-range photogrammetry could be a valid and reliable technique for creating 3D models of museum artifacts (Nicolae et al., 2014). Marshall et al. presented a novel object scanner for photogrammetry image acquisition across large museum collections (Marshall et al., 2019). Bucchi et al. described a comprehensive guideline to improve 3D model quality (Bucchi et al., 2020). With the wide application of this technique, close-range photogrammetry is also used in the digitalization of immovable heritage assets. The increased use is most evident with examples in buildings and monuments (Yastikli, 2007, Yilmaz et al., 2007, Yilmaz et al., 2008). Here, a great deal of interest has been attracted to 3D reconstruction technologies that combine close-range photogrammetry with Structure from Motion (SfM) and enable the creation of dense point clouds from images captured using conventional digital cameras (McCarthy, 2014, Giuliano, 2014, Kasapakis et al., 2019). Due to the development of SfM and multi-view stereo (MVS) algorithms, aerial photogrammetry by UAVs achieves great success in the easy, intuitive, and fast construction of large-scale scenes. It has become an important approach for the 3D documentation of archaeological excavations, landscapes, and relatively large sites (Campana, 2017, Pepe et al., 2022). Moreover, digitalizing immovable heritage assets is usually combined with traditional surveying and mapping methods. Developments in photogrammetry and terrestrial laser scanning workflows are creating increasingly comprehensive and accurate 3D colored point clouds (Hassan and Fritsch, 2019). The 3D documentation of large-scale and complex structure objects is often accomplished today by combining photogrammetric processes and terrestrial laser scanners (Moussa et al., 2013).

However, the lack of surface texture of some heritage assets makes it difficult to establish high-precision 3D models by simply using images. To solve the problem, researchers usually use structured light scanners (SLS) to collect geometric models and map textures (Gomes et al., 2014). Representative examples include Stanford University’s Michelangelo Project (Levoy et al., 2000), IBM’s Pieta Project (Bernardini et al., 2002), and Columbia University’s Beauvais Cathedral Project (Allen et al., 2003), which greatly impacted the field. With the popularization and cost reduction of structured light measurement equipment, it has become one of the most important tools in digitalizing heritage assets (Akça, 2012). Terrestrial laser scanners (TLS) are also heavily used in large-scale heritage site mapping because of their capability to acquire a lot of information with a high degree of detail in a relatively short time (millimeter accuracy of over 500,000 points per second) (Castagnetti et al., 2012). After a decade of continuous development, TLS workflows have reached a high level of automation in cultural heritage surveys (Vacca et al., 2012, Pritchard et al., 2017, Kushwaha et al., 2020). At the same time, the emergence of various studies on the configuration of terrestrial laser scanning power

* These authors contributed equally to this work.
† Corresponding author

https://doi.org/10.5194/isprs-annals-X-2-2024-231-2024

supply, simplified models, components, strengths, corrections, etc., makes this technology more convenient in the application of cultural heritage. Furthermore, researchers have made unprecedented discoveries in archaeological applications by placing laser scanning on airplanes or drones (Chase et al., 2012, Evans et al., 2013, Prümers et al., 2022). The above laser scanning techniques enable the collection of accurate and detailed 3D point clouds. However, due to specific requirements in different reconstruction projects, none of the scanners are superior to the others (Moussa, n.d.). Therefore, a fusion methodology of multiple data and sensors is required to allow the large-scale acquisition of heritage sites with locally complex structures (Shao et al., 2019).

Although there have been great technological developments in traditional photogrammetry, some methods that integrate artificial intelligence and photogrammetry have developed fast and demonstrated powerful model-building capabilities, such as NeRF and other technologies. However, there are still many problems before these technologies are practical (Mildenhall et al., 2021, Wang et al., 2021, Kerbl et al., 2023). The high-precision digitalization of large-scale grottoes, especially those with extremely rich details, has not been well solved so far.

The digitalization accuracy of grottoes is required to realize 3D printing and reproduction without distortion. Therefore, there are specific requirements for geometric accuracy and texture resolution. For example, geometric accuracy should be better than millimeters, texture resolution should be better than 150 DPI, and it is best to reach 300 DPI. Current technical means can achieve this precision in small scenes. However, limited by the accuracy and processing techniques of existing sensors, it becomes increasingly difficult to ensure overall geometric accuracy and the geometric accuracy and texture resolution of local details when the height of the grottoes is more than 10 meters. The single point accuracy of TLS is about 2-3 mm, which makes it difficult to achieve local millimeter accuracy in data expression. However, 3D reconstruction methods using photogrammetry to ensure the requirements of local details and texture resolution will make aerial triangulation unstable and cannot guarantee overall accuracy.

This paper studies the high-precision 3D digitalization technology route of large-scale cultural heritage. Taking Cave No. 13 in Yungang Grottoes as an example, we study the entire workflow from data acquisition to data processing, use TLS as the whole control to ensure the overall accuracy of the scene and use photogrammetric modeling to meet the geometric accuracy requirements of local details in the digitalization of heritage assets.

2. Yungang Grottoes Cave No. 13

Yungang Grottoes is located on the southern part of Wuzhou Mountain, about 16 kilometers west of Datong City, Shanxi Province. The 254 extant grottoes consist of 45 major grottoes, more than 1,100 Buddhist niches, and more than 51,000 sculptures. Representing the highest level of world carved art in the fifth century, Yungang Grottoes is considered exemplary Chinese Buddhist artwork at its finest. It is known as the three major grottoes in China, with Mogao Grottoes in Dunhuang and Longmen Grottoes in Luoyang, and one of the world’s three major carved art treasures, with Ajanta Grottoes in India and Bamiyan Grottoes in Afghanistan. In December 2001, Yungang Grottoes was listed as a UNESCO World Heritage Site and became one of the nation’s earliest 5A tourist areas in May 2007.

Cave No. 13 is a typical cave in the Yungang Grottoes group, located in the middle of the Grottoes group in the west, about 11 meters wide, 9 meters deep, and 13.6 meters high. There is a main Buddha statue, with over 2,900 body sculptures carved on the wall. The geometric details are rich, and many colorful paintings dating back 1,500 years are retained, which poses a great challenge to photogrammetric 3D reconstruction (Figure 1).

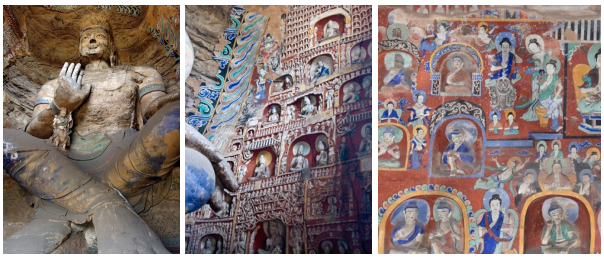


Figure 1. The main Buddha statue of Yungang Grottoes Cave No. 13 and sculptures carved on the wall.

3. Method

3.1 Workflow

Cave No. 13 in Yungang Grottoes has a large volume and many sculptures and murals, with complex structures and compositions attached to its wall. Therefore, the challenge of digital modeling technology is high. The whole technology route needs to consider two aspects: data acquisition and data processing.

In Figure 2, we propose a workflow to solve several technical problems in high-precision digital 3D reconstruction of such large-scale and complex structure grottoes. The workflow mainly includes large-space laser scanning, image acquisition, large-scale integral aerial triangulation, geometric reconstruction, color processing, and other key technical links. In this workflow, we control data through large-space laser scanning, which provides a unified spatial reference for the photogrammetric solution. To obtain high-resolution original images of the entire cave, we propose to build scaffolding and shoot close to images in layers during the data acquisition stage and carry out an aerial triangulation verification on image data of each scaffold layer to ensure the integrity of the image acquisition. The aerial triangulation of such large and completely disordered images is also challenging. We propose the aerial triangulation strategy of stratification first and then overall, and the results of stratified aerial triangulation provide initial values for the overall aerial triangulation, thus ensuring the execution of overall aerial triangulation. This workflow takes full advantage of multiple sensors, realizes the rapid processing and fusion of multi-sensor data, and the reconstruction and rapid visualization of 3D models of large-scale and complex structure grottoes. We verify the feasibility and reliability of this overall workflow in Yungang Grottoes.

Due to the large volume of the entire cave, shooting from a close distance is necessary to ensure the accuracy of image texture resolution and geometric reconstruction. Therefore, the collection interface will be divided into seven layers in the internal structure, and scaffolding will be built every 2 meters. Data acquisition is collected from the upper layer, and one layer is collected and then removed. The data acquisition process includes

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-2-2024-231-2024 | © Author(s) 2024. CC BY 4.0 License.

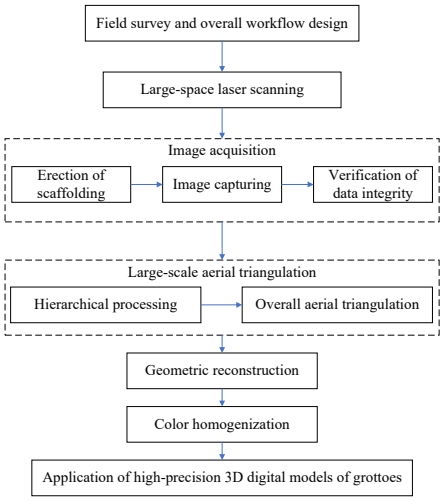


Figure 2. Overall workflow.

terrestrial laser scanning and two camera control contents. As shown in Figure 3, the scaffold working platform is set up at 2 meters, 4 meters, 6 meters, 8 meters, 10 meters, and 12 meters above the ground to form seven layers of the working platform, including the ground layer.

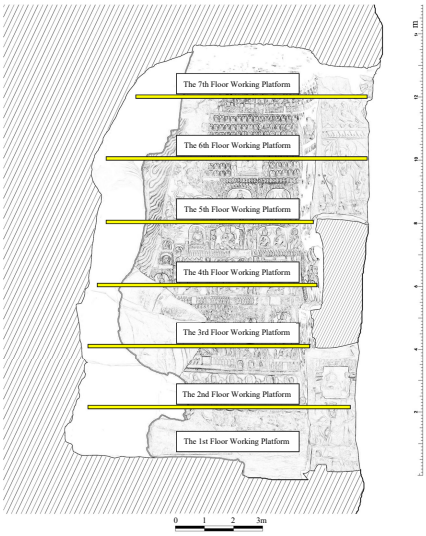


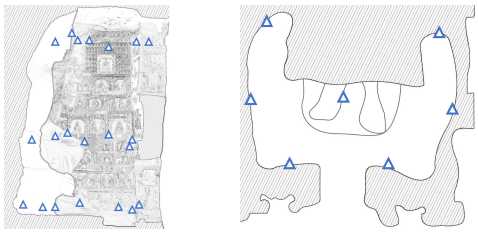
Figure 3. Scaffolding working platform in Yungang Grottoes Cave No. 13.

3.2 Data Acquisition

3.2.1 Control network layout and measurement. The data acquisition of Yungang Grottoes has the characteristics of a large volume of grottoes, fine data collection, long duration of data collection process, and multiple types of equipment used for data collection. During data acquisition, the overall control network of the cave must be established to incorporate such diverse and extensive data into a unified spatial reference. Two aspects should be considered in the layout of control points of the control network: (1) Control points should be evenly distributed on the inner surface of the cave, and the uniformity of distribution should be taken into account in both vertical and horizontal directions; (2) Since the surface of the cave is a precious heritage asset and no equipment is allowed to touch it, it is impossible to use a fixed target on the wall commonly used in building data acquisition as the control point, but natural feature points inside the cave should be selected as control points, and the collection of the cave has both geometric and color data. Therefore, it is necessary to select points where geometric and texture features are obvious and can be observed from multiple views. As shown in Figure 4, the decorative part of the canopy in the left figure has a sharp geometric feature, which is easy to locate, and the color information also shows an obvious jump there. The eyes of the Buddha statue in the right figure also have obvious features in geometry and texture, so they are suitable as candidates for feature points.

Figure 4. Diagram of points suitable for use as control points.

Given the volume size of Cave No. 13 in Yungang Grottoes and the characteristics of Buddha statues and walls, five points with obvious geometric and texture characteristics should be selected as control points at three different cave heights. A total of 15 control points constitute the control network of Cave No. 13 in Yungang Grottoes. The vertical distribution of control points in the control network is shown in Figure 5(a). Control points will be selected at 1 meter, 6 meters, and 11 meters above the ground. The height of control points at each height level can float within a range of 2 meters. The horizontal distribution of control points in the control network is shown in Figure 5(b). Seven appropriate control points are selected on the main Buddha statue and the wall at each height level.



(a) Vertical distribution. (b) Horizontal distribution.

Figure 5. Distribution of control points.

In control network measurement, considering the restriction on the absolute value of coordinate values in fine 3D digital reconstruction and the preservation of more significant digits for the reconstructed geometric model, the control network of Cave No. 13 in Yungang Grottoes established by this method is a local relative coordinate system, and the coordinate origin is selected in the cave. If the follow-up work needs to be combined with the national standard geodetic coordinate system, it is only necessary to introduce the control point coordinates of the known geodetic coordinate system into the control network and calculate the rotation matrix.

We exploit a combined total station observation and laser scanning method to measure the high-precision control network. The measurement method is: (1) Control point coordinates measurement using a total station. Since control points in the cave

https://doi.org/10.5194/isprs-annals-X-2-2024-231-2024

can only be measured by the prism-free mode of the total station, it is necessary to observe control point coordinates in the control network from multiple observation sites and adjust coordinate values to improve accuracy. (2) Laser scanning combined solution. To further improve the precision of the control network, the coordinates observed by the total station and the coordinates fitted by the laser scanner are combined to solve the adjustment after laser scanning data collection is completed.

3.2.2 Geometric information collection of the entire cave. Based on the overall control network, this method adopts a 3D scanning instrument to obtain geometric data of the cave. Cave No. 13 in Yungang Grottoes is a roofed cave environment, so selecting a large-space scanner that can scan to the top in the vertical direction is necessary. Some areas in the cave are narrow, and to obtain more comprehensive data, the closest scanning distance of the scanner should be as short as possible. After considering these factors and comparing mainstream 3D laser scanners, the Faro Focus3D X330 3D laser scanner is adopted for data acquisition in this project. The principle of large-space scanner station layout is to face the scanned object as far as possible. The layout scheme of the large space scanner station on the working platform of each layer designed by this method for Cave No. 13 in Yungang Grottoes is shown in Figure 6. Each layer of the working platform needs to scan about eight stations of 3D laser point cloud data. The specific location and number of stations will be adjusted according to the site situation and different structures of each layer to leave no blind angle. As shown in Figure 6, this layout ensures that surface coverage exceeds 90% and overlap between stations is expected to exceed 60%.

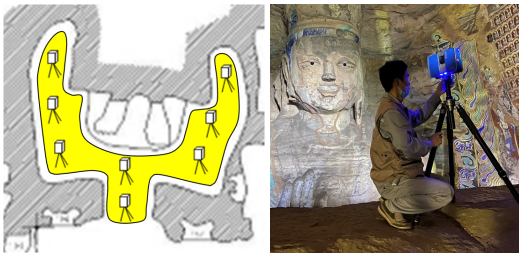


Figure 6. Large-space scanner station layout in Yungang Grottoes Cave No. 13.

3.2.3 Texture information collection of the entire cave. In image acquisition of objects with complex shapes, the shooting angle should be comprehensive, and the overlap between adjacent images should be more than 50% to ensure that the subsequent 3D reconstruction can avoid holes.

The images are captured on a Canon EOS-5Ds Mark IV digital camera with 50 mm and 100 mm prime lenses. According to the digitalization accuracy requirements of the cave, the pixel density of surface texture should not be less than 34.88 million pixels per square meter (150 DPI). For the image acquisition of Cave No. 13 in Yungang Grottoes, the calculated shooting distance is 1.3 m for 50 mm and 2.6 m for 100 mm according to resolution, camera sensor size, and lens focal length. In practice, we shoot 300 DPI images using a 50 mm lens with a shooting distance of 1.3 m and a 100 mm lens with a shooting distance of 2.6 m. According to this shooting method, the ground sample distance (GSD) of the surface texture captured is 2 mm.

During the acquisition process, the color card information is also collected to correct the color of the grotto texture data (Figure 7). Figure 8 is a comparison of images before and after color correction.

Figure 7. Photography working scene and used color cards.

Figure 8. The color-corrected image using color cards. The left is before correction, while the right is after correction.

3.3 Data Processing

There are three main problems in 3D reconstruction of large scene multi-view images: (1) scale information, (2) aerial triangulation solution, and (3) color consistency. Since photogrammetry follows a triangulation measurement principle, additional ground truth must be introduced to determine the absolute scale. In this paper, this is solved by utilizing the high-precision control network measured by total station observation and laser scanning. Another issue is that the amount of image data collected during the 3D reconstruction of a large-scale scene is often huge. However, existing reconstruction methods all adopt the idea of dividing and ruling, bringing some problems. On the one hand, the method of partition and block is used to solve aerial triangulation of image data, and there will be a problem of uneven resolution accuracy among each block, which will lead to the phenomenon of stratification at the edge of each block in the reconstructed geometric model. On the other hand, for large-scale 3D scene reconstruction, due to the influence of illumination changes and other factors during acquisition, acquired image data will have inconsistent brightness, eventually leading to large color differences in generated results.

3.3.1 Distributed aerial triangulation based on alternating parameter solution. In this paper, the distributed global solution technique of aerial triangulation based on alternating parameters is used to realize error compensation and optimization between partitioned data, which avoids reliability reduction caused by the partitioned solution and greatly improves overall efficiency.

Bundle adjustment is an automatic method to obtain 3D information of the photographed object from images. It takes the coordinates of image points as observed values and, at the same time, calculates camera parameters and the coordinates of object points. Essentially, it is a sparse nonlinear least squares problem with a special structure, which can be solved efficiently using its sparse structure.

For large-scale photogrammetric data (tens of thousands or hundreds of thousands of images), the images are usually processed by computer clusters and divided into blocks to reduce data processing time. Each block’s camera parameters and object coordinates are recovered independently by different computing

nodes. Then, the coordinate system of each block data is unified, and an overall bundle adjustment is carried out. However, the geometric accuracy of each block is not uniform, and a globally optimal solution cannot be obtained. This paper uses an ADMM method to achieve a large-scale global solution. Generally, unknowns in the bundle method include cameras' internal and external parameters and an object 3D point set observed by identical point pairs. A camera has the same intrinsic parameters, while the camera has different extrinsic parameters for each image taken. Then, based on the collinear equation, our objective function is:

$$f(I,E,P)=\sum_{m=1}^{n}\sum_{p_j\in S_m}\left\|\Pi\left(i_m,e_{lm},p_j\right)-q_{mj}\right\|, \tag{1}$$

where I represents the set of internal parameters for each camera, E represents the set of external parameters corresponding to each camera position, P is the set of object points, i_m is the internal parameter of the m_{th} camera, e_{lm} is the l_{th} image with the m_{th} internal parameter, p_j is the j_{th} object point, q_{mj} is the image coordinate of the object point on the image with the m_{th} internal parameter, and Π is the projection model.

The alternating direction multiplier method (ADMM) is important for solving separable convex optimization problems. Due to fast processing speed and good convergence performance, the ADMM algorithm has been widely used in statistical learning, machine learning, and other fields. The ADMM is generally used to solve the following convex optimization problems:

$$\begin{aligned}&\text{minimize } f(x)+g(y)\\&\text{subject to } Ax+By=c.\end{aligned} \tag{2}$$

According to the Lagrange multiplier method, we can get:

$$\begin{aligned}&L_p(x,y,\lambda)=f(x)+g(y)+\lambda^T(Ax+By-c)\\&+\tfrac{\rho}{2}\|Ax+By-c\|_2^2,\rho>c,\end{aligned} \tag{3}$$

where λ is the Lagrange multiplier, ρ is the penalty coefficient, then the iterative solution process of the ADMM algorithm is as follows:

$$\begin{aligned}x^{k+1}&=\arg\min L_p(x,y,\lambda)\\y^{k+1}&=\arg\min L_p(x,y,\lambda)\\\lambda&=\lambda^k+\rho\left(Ax^{k+1}+By^{k+1}-c\right).\end{aligned} \tag{4}$$

In the actual solving process, alternating iteration is carried out. Then, we will solve the overall objective function as follows:

$$\begin{aligned}&\min\textstyle\sum_{i=1}^{n}f_i(x_i)\\&\text{subject to } x_i=y,i=1,\ldots,n.\end{aligned} \tag{5}$$

Using the ADMM algorithm, we can obtain the following iterative process:

$$\begin{aligned}x_i^{t+1}&=\arg\min\left(f_i(x_i)+\left(\lambda_i^t\right)^T\left(x_i-y^t\right)+\tfrac{\rho}{2}\left\|x_i-y^t\right\|_2^2\right)\\y^{t+1}&=\tfrac{1}{n}\textstyle\sum_{i=1}^{n}x_i^{t+1}\\\lambda_i^{t+1}&=\lambda_i^t+\rho\left(x_i^{k+1}-y^{k+1}\right).\end{aligned} \tag{6}$$

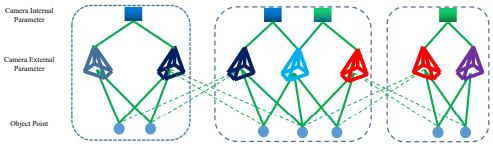


Figure 9. Schematic diagram of ADMM-based global consistency solving for cameras.

As shown in Figure 9, the global camera parameter consistency solution method divides the whole data region into K subblocks. Objects with the same color between blocks represent the same parameters, with the common camera parameter and the same external parameters (multiple subblocks of the same image). Therefore, to keep the cameras' internal and external parameters corresponding to the same image between blocks consistent, the original collinear equation objective function is modified as follows:

$$\begin{aligned}&\min\textstyle\sum_{k=1}^{K}f(I_k,E_k,P_k),i_m^k=i_m,k=1,\ldots,K\\&\text{subject to } e_m^k=e_m,k=1,\ldots,K.\end{aligned} \tag{7}$$

According to the ADMM algorithm, we can solve the above problems through iteration:

$$\begin{aligned}&(I_k,E_k,P_k)^{t+1}=\min f(I_k,E_k,P_k)+g(f(I_k,E_k,P_k))\\&i_m^{t+1}=\textstyle\sum\left(e_i^k\right)^{t+1}/n_m,e_m^{t+1}=\sum\left(d_i^k\right)^{t+1}/m_l\\&\left(\tilde{i}_m^k\right)^{t+1}=\left(\tilde{i}_m\right)^t+\left(i_m^k\right)^{t+1}-i_m^{t+1}\\&\left(\tilde{e}_m^k\right)^{t+1}=\left(\tilde{e}_m\right)^t+\left(e_m^k\right)^{t+1}-e_m^{t+1},\end{aligned} \tag{8}$$

where $f(I_k,E_k,P_k)$ is the objective function of the k_{th} block, and $g(f(I_k,E_k,P_k))$ is:

$$\begin{aligned}&g(f(I_k,E_k,P_k))=\\&\textstyle\sum_{i_m^k\in I_K}\left\|i_m^k-i_m^t+\left(\tilde{i}_m^k\right)^t\right\|^2_{\sum_i}\\&+\textstyle\sum_{e_n^k\in E_K}\left\|e_n^k-e_n^t+\left(\tilde{e}_n^k\right)^t\right\|^2_{\sum_e}\\&+\rho_p\textstyle\sum_{p_j\in P_K}\left\|p_j-p_j^t\right\|_2^2.\end{aligned} \tag{9}$$

In the process of solving, it is necessary to divide tiles as evenly as possible, and synchronization of each subblock is calculated in time. At the same time, the initial value of the Lagrange multiplier is more important in the solving process because the dimensions of the cameras' internal and external parameters are inconsistent.

3.3.2 Aerial triangulation based on control network. The SfM reconstruction process delivers accurate camera orientations and sparse point clouds, but initially in an arbitrary model

space. We fill gaps in image-based 3D reconstruction pipelines and retrieve scale information using the control network measured with total station observation and laser scanning. Since no equipment is allowed to touch heritage assets, placing fixed targets is impossible, but natural feature points should be selected as control points. To build a correlation between 2D images and 3D control points, we first extract obvious feature points from the intensity map of the laser field. Each pixel in the 2D intensity map corresponds to one 3D laser scanning point, and the number of pixels results from the defined scan resolution. Then, 2D-to-2D correspondences between photographed images and intensity maps can be established, allowing an implicit determination of 3D-to-3D correspondences between sparse point clouds and laser scanning data. These correspondences are applied to the SfM output and provide absolute camera orientations related to the laser data by introducing the scale information to the bundle.

3.3.3 Color homogenization based on global adjustment. For 3D reconstruction of large-scale scenes, it is generally necessary to divide a large range of scenes into multiple tiles for reconstruction. However, in the process of oblique image acquisition, due to the influence of illumination changes and other factors, collected image data will have inconsistent brightness, leading to the appearance of color difference seams in the surface texture of the generated 3D scene. Although the method proposed by Waechter et al. (Waechter et al., 2014) can eliminate texture color difference seams inside a single tile, there will also be inconsistent brightness between tiles. Therefore, this paper designs a color homogenization method for 3D models across tiles based on global adjustment. The basic idea of this method is to homogenize all image data globally before texture mapping to ensure the color consistency of the generated 3D tile model as much as possible. After texture mapping, 3D tiles of the scene are uniformly colored to eliminate the problem of inconsistent brightness at the seams of tiles.

The core idea of the method is to establish a linear equation of color differences of all tiles in measuring areas by using color differences of the 3D real scene in overlapping areas. The linear equation is solved as a whole, and the change of 3D color of each tile is obtained to achieve the effect of overall uniform color.

4. Result and Discussion

4.1 Result

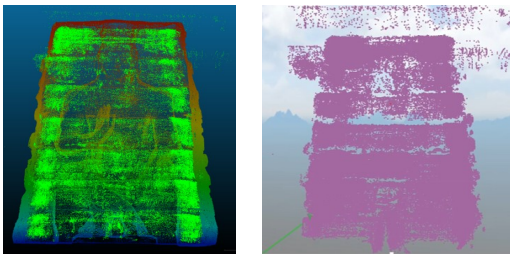
We obtain 2.1 TB laser scanning point clouds (433 stations) and 283,400 original images in the data acquisition phase. After data quality inspection and precision assessment, the median error of feature point spacing of point clouds is less than 5 mm, the maximum point spacing is less than 3 mm, surface coverage is greater than 80%, and point cloud overlap between adjacent scanning stations is greater than 30%.

In the data processing stage, we develop Get3d Cluster, an automatic photogrammetric modeling software supporting a parallel 3D modeling system with millions of images and thousands of CPUs. The software integrates a stable and reliable distributed aerial triangulation method for solving alternating parameters and an integrated color homogenization method for 3D models across tiles based on global adjustment. To deal with such a large amount of data solving, we adopt the strategy of a large number of GPU server clusters. We use 120 server nodes for cluster computing. After about 85 days of data processing, the overall aerial triangulation and 3D reconstruction of Cave No. 13 in Yungang Grottoes are realized.

To establish the control point of the image, we use the intensity map of the laser field to find some obvious feature points, which is used to build a correlation with the image collected by the camera. The final statistics of aerial triangulation access images are 239,700, and the access rate reaches 84.58%. By calculating the random sampling back-projection error of aerial triangulation, the error of our aerial triangulation results is better than 0.5mm. The results of aerial triangulation are shown in Figure 10.

Through geometric reconstruction after aerial triangulation and color homogenization across tiles based on global adjustment, we obtain the entire 3D model of Cave No. 13 in Yungang Grottoes. The model is in OBJ and OSGB formats, the data volume is 500 GB, a total of 880 million triangles, and it is divided into 5,140 tile storage. We exploit the visualization strategy of model reconstruction in multi-level details. A levels-of-detail model is constructed to improve loading and browsing efficiency. Figure 11 shows the 3D model of the entire cave. Figure 12(a) shows a partial mesh model.

It is estimated that the texture resolution of the 3D model we obtained reaches 300 DPI, the back-projection error of control points is less than 5 pixels, and the misalignment between model outlines is less than 5 pixels. Figures 12(b), 12(c), and 12(d) show details of the model, including the partial model, the shrine, and the dome.



(a) Connecting points of aerial triangulation.

(b) Each point corresponds to one photo, a total of 239,000 photos.

Figure 10. Results of aerial triangulation.

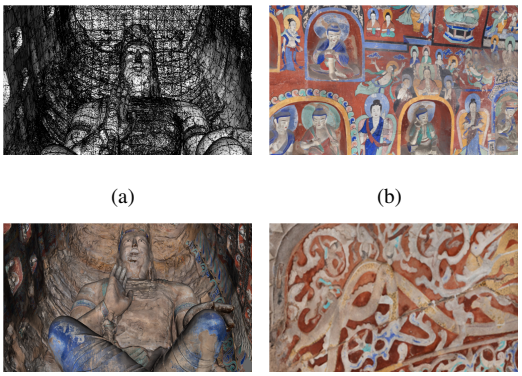
4.2 Application Using 3D Models

4.2.1 3D display and information superposition. The successful construction of high-precision 3D digitalization results of Cave No. 13 in Yungang Grottoes can break through display requirements and fidelity effects that traditional protection methods cannot achieve. We develop a digital display system for Cave No. 13 in Yungang Grottoes to better display and fully use this precious digital model. Based on the 3D grottoes display system, users can use 3D GIS to overlay data and establish information association so that 3D models can be integrated with humanistic research information (Figure 13).

4.2.2 3D disease information labeling. With a high-precision 3D model, users can add, modify, and query diseases on the 3D model through visual interpretation. Due to the high geometric accuracy of the entire model, diseases plotted can be quantitatively analyzed and counted, which greatly improves the efficiency of cultural heritage protection and restoration (Figure 14).

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-2-2024-231-2024 | © Author(s) 2024. CC BY 4.0 License.

Figure 11. The 3D model of the entire Yungang Grottoes Cave No. 13.



(a) (b)

(c) (d)

Figure 12. Results of 3D reconstruction.

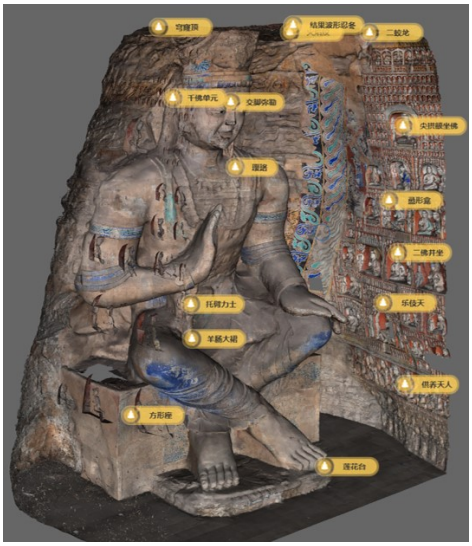


Figure 13. Annotation and overlay of information based on the 3D model.

4.3 Future Work

4.3.1 1:1 high-precision reproduction based on 3D printing. In recent years, Yungang Research Institute has successfully reproduced Caves No. 3, No. 12, and No. 18. The high-precision digitalization results of Cave No. 13, constructed by this project, provide the most important data for reproducing the entire cave. Through 3D printing technology, the digitalization results can be transformed into physical grottoes and popularized to the public.

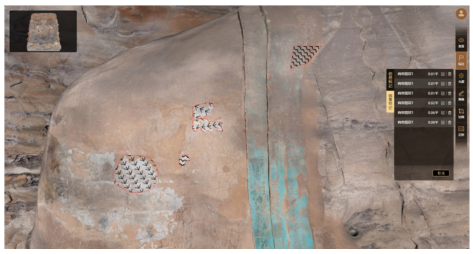


Figure 14. Disease annotation based on the 3D model.

4.3.2 Monitoring and analysis of grotto changes based on time series. The serialization information of grottoes is collected at regular intervals to form a complete 4D database for change monitoring. Combined with synchronous records and archives of other information, such as the environment, it can be used for professional analysis. The detailed grotto database will provide scientific researchers with data that can be easily retrieved, accurately measured, and used for global spatial analysis based on environmental spatial data. Data are classified according to spatial location, subject, time, and category in the database, which is convenient for efficient retrieval.

5. Conclusion

In this paper, we propose a holistic solution to high-precision digitalization of large-scale grottoes. Laser scanning is used for overall control, and photogrammetry is used to collect a large number of high-precision images. A parallel ADMM algorithm solves overall aerial triangulation and high-precision scene reconstruction of large-scale images. Taking Cave No. 13 in Yungang Grottoes as an example, high-precision 3D cave modeling with more than 280,000 images is realized. This method can take advantage of the overall accuracy of terrestrial laser scanners and the high resolution and expressiveness of photogrammetry in local details. The project case in this paper provides a good reference for high-precision digital modeling of large-scale heritage assets.

References

Akça, M. D., 2012. 3D modeling of cultural heritage objects with a structured light system. *Mediterranean Archaeology and Archaeometry*, 12(1), 139–152.

Allen, P. K., Troccoli, A., Smith, B., Stamos, I., Murray, S., 2003. The beauvais cathedral project. *Conference on Computer Vision and Pattern Recognition Workshop 2003*, 1, 10–10.

Bernardini, F., Rushmeier, H., Martin, I. M., Mittleman, J., Taubin, G., 2002. Building a digital model of Michelangelo’s Florentine Pieta. *IEEE Computer Graphics and Applications*, 22(1), 59–67.

Bucchi, A., Luengo, J., Fuentes, R., Arellano-Villalón, M., Lorenzo, C., 2020. Recommendations for improving photo quality in close range photogrammetry, exemplified in hand bones of chimpanzees and gorillas. *International Journal of Morphology*, 38(2), 348–355.

Campana, S., 2017. Drones in archaeology. State-of-the-art and future perspectives. *Archaeological Prospection*, 24(4), 275–296.

Castagnetti, C., Bertacchini, E., Capra, A., Dubbini, M. et al., 2012. Terrestrial laser scanning for preserving cultural heritage: analysis of geometric anomalies for ancient structures. *Proceedings of FIG Working Week 2012*, FIG Federation International des Geometres, 1–13.

Chase, A. F., Chase, D. Z., Fisher, C. T., Leisz, S. J., Weishampel, J. F., 2012. Geospatial revolution and remote sensing LiDAR in Mesoamerican archaeology. *Proceedings of the National Academy of Sciences*, 109(32), 12916–12921.

De Fino, M., Galantucci, R. A., Fatiguso, F., 2023. Condition assessment of heritage buildings via photogrammetry: a scoping review from the perspective of decision makers. *Heritage*, 6(11), 7031–7066.

Evans, D. H., Fletcher, R. J., Pottier, C., Chevance, J.-B., Soutif, D., Tan, B. S., Im, S., Ea, D., Tin, T., Kim, S. et al., 2013. Uncovering archaeological landscapes at Angkor using LiDAR. *Proceedings of the National Academy of Sciences*, 110(31), 12595–12600.

Giuliano, M., 2014. Cultural heritage: An example of graphical documentation with automated photogrammetric systems. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40(5), 251–255.

Gomes, L., Bellon, O. R. P., Silva, L., 2014. 3D reconstruction methods for digital preservation of cultural heritage: A survey. *Pattern Recognition Letters*, 50, 3–14.

Hassan, A. T., Fritsch, D., 2019. Integration of laser scanning and photogrammetry in 3D/4D cultural heritage preservation-a review. *International Journal of Applied*, 9(4), 16.

Jebur, A. K., 2022. The techniques of cultural heritage: literature review. *Saudi J. Civ. Eng*, 6(4), 108–114.

Kasapakis, V., Gavalas, D., Dzardanova, E., 2019. 3d modelling through photogrammetry in cultural heritage.

Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G., 2023. 3D gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4), 1–14.

Kushwaha, S., Dayal, K. R., Sachchidanand, Raghavendra, S., Pande, H., Tiwari, P. S., Agrawal, S., Srivastava, S., 2020. 3d digital documentation of a cultural heritage site using terrestrial laser scanner—a case study. *Applications of Geomatics in Civil Engineering: Select Proceedings of ICGCE 2018*, 33, 49–58.

Levoy, M., Pulli, K., Curless, B., Rusinkiewicz, S., Koller, D., Pereira, L., Ginzton, M., Anderson, S., Davis, J., Ginsberg, J. et al., 2000. The digital michelangelo project: 3d scanning of large statues. *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, 131–144.

Marshall, M., Johnson, A., Summerskill, S., Baird, Q., Esteban, E., 2019. Automating photogrammetry for the 3D digitisation of small artefact collections. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42(2), 751–757.

McCarthy, J., 2014. Multi-image photogrammetry as a practical tool for cultural heritage survey and community engagement. *Journal of Archaeological Science*, 43, 175–185.

Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., Ng, R., 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1), 99–106.

Moussa, W., n.d. Integration of digital photogrammetry and terrestrial laser scanning for cultural heritage data recording. PhD thesis, Stuttgart, Universität Stuttgart, 2014.

Moussa, W., Wenzel, K., Rothermel, M., Abdel-Wahab, M., Fritsch, D., 2013. Complementing TLS point clouds by dense image matching. *International Journal of Heritage in the Digital Era*, 2(3), 453–470.

Nicolae, C., Nocerino, E., Menna, F., Remondino, F., 2014. Photogrammetry applied to problematic artefacts. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40(5), 451–456.

Pepe, M., Alfio, V. S., Costantino, D., 2022. UAV platforms and the SfM-MVS approach in the 3D surveys and modelling: A review in the cultural heritage field. *Applied Sciences*, 12(24), 12886.

Pritchard, D., Sperner, J., Hoepner, S., Tenschert, R., 2017. Terrestrial laser scanning for heritage conservation: The Cologne Cathedral documentation project. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4(2), 213–220.

Prümers, H., Betancourt, C. J., Iriarte, J., Robinson, M., Schaich, M., 2022. LiDAR reveals pre-Hispanic low-density urbanism in the Bolivian Amazon. *Nature*, 606(7913), 325–328.

Shao, J., Zhang, W., Mellado, N., Grussenmeyer, P., Li, R., Chen, Y., Wan, P., Zhang, X., Cai, S., 2019. Automated markerless registration of point clouds from TLS and structured light scanner for heritage documentation. *Journal of Cultural Heritage*, 35, 16–24.

Vacca, G., Deidda, M., DESSI, A., Marras, M. et al., 2012. Laser scanner survey to cultural heritage conservation and restoration. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 39(B5), 589–594.

Waechter, M., Moehrle, N., Goesele, M., 2014. Let there be color! large-scale texturing of 3d reconstructions. *European Conference on Computer Vision 2014*, 8693, 836–850.

Wang, P., Liu, L., Liu, Y., Theobalt, C., Komura, T., Wang, W., 2021. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. *Advances in Neural Information Processing Systems*, 34, 27171–27183.

Yastikli, N., 2007. Documentation of cultural heritage using digital photogrammetry and laser scanning. *Journal of Cultural heritage*, 8(4), 423–427.

Yilmaz, H. M., Yakar, M., Gulec, S. A., Dulgerler, O. N., 2007. Importance of digital close-range photogrammetry in documentation of cultural heritage. *Journal of Cultural Heritage*, 8(4), 428–433.

Yilmaz, H., Yakar, M., Yildiz, F., 2008. Documentation of historical caravansaries by digital close range photogrammetry. *Automation in Construction*, 17(4), 489–498.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-2-2024-231-2024 | © Author(s) 2024. CC BY 4.0 License.