Available Online: 31 July 2026 DOI: 10.54254/2977-3903/2026.35765

Research on digital modeling of ancient building appearances based on UAV three-dimensional reconstruction

Yuechen Yin

Yali High School International Department, Changsha, China

1099068134@qq.com

Abstract. A digital exterior-modeling method based on Unmanned Aerial Vehicle (UAV) oblique photogrammetry and point cloud processing is proposed. The difficulty encountered by conventional surveying methods in comprehensively acquiring spatial information from elevated and occluded areas of historic buildings is addressed by this method. The Jiuyun Fangding monument was selected as the study object. Images were acquired using a multi-altitude, multi-angle, layered circumferential flight strategy. A Three-Dimensional (3D) model was then reconstructed through feature matching, camera-pose estimation, multi-view stereo matching, point cloud registration, and texture mapping. Four representative dimensions were selected, and the measurements obtained from the reconstructed model were compared with field measurements. The results indicated that: (1) the absolute errors of the four dimensions ranged from 0.006 to 0.039 m, with the maximum spacing between the load-bearing columns exhibiting the largest absolute error of 0.039 m; (2) the relative errors ranged from 0.63% to 1.81%, with the width of the load-bearing column exhibiting the largest relative error of 1.81%; and (3) the reconstructed model provided a relatively complete representation of the overall architectural form and the principal structural components. The proposed method can therefore provide technical support for the digital documentation, 3D visualization, and dimensional verification of historic buildings.

Keywords: UAV oblique photography, ancient buildings, 3D reconstruction, point cloud processing, accuracy validation

1. Introduction

In recent years, the performance of UAVs has continuously improved, while their operational costs have gradually decreased, owing to advances in flight control, satellite positioning, and intelligent sensing technologies. UAVs have consequently been widely applied in industrial production and everyday activities [1-3]. Because of their compact size, high maneuverability, and ease of deployment, areas that are difficult for personnel or conventional equipment to access can be reached rapidly by the UAVs. When equipped with different types of sensors, UAVs can be used for information acquisition, target monitoring, and related tasks [4-6]. Moreover, advances in automated route planning, real-time data transmission, and intelligent processing have further improved the efficiency and stability of UAV operations [7-9]. Thus, UAVs have gradually evolved from individual flight devices into integrated technological platforms and have become important tools for promoting digitalization and intelligent development in various fields.

Systematic investigations of the 3D digital documentation of historic buildings and cultural heritage sites have been conducted worldwide [10-12]. Wang et al. [13] reported that manual measurement using tape measures and distance meters is labor-intensive and inefficient. Such approaches also make it difficult to obtain dimensional information from elevated or concealed structural components. Lovrincevic et al. [14] noted that instruments such as total stations and theodolites primarily collect data at discrete points and are therefore constrained by instrument placement and line-of-sight conditions. Consequently, the spatial forms of complex components cannot be represented comprehensively using these instruments. Jia et al. [15] reported that terrestrial 3D laser scanning can provide high measurement accuracy. However, the equipment is expensive, and point cloud registration across multiple stations is cumbersome. In addition, data may be missing from occluded areas. Overall, existing ground-based surveying approaches cannot simultaneously satisfy the requirements of efficiency, cost-effectiveness, and data completeness and therefore remain inadequate for the comprehensive digital documentation of complex historic buildings.

To overcome the limitations of conventional ground-based approaches in terms of operational efficiency, visibility, and coverage of complex areas, UAV photogrammetry provides an alternative technical solution. Wang et al. [16] found that, compared with conventional methods, UAVs can improve data-acquisition efficiency while maintaining the measurement accuracy. Jia et al. [17] demonstrated that multi-angle UAV images can reduce the data loss caused by line-of-sight occlusion and improve coverage of geometrically complex targets. Lu et al. [18] reported that UAV photogrammetry does not require direct contact with the surveyed object, thereby reducing the difficulty and cost of field operations. Furthermore, the spatial form and the surface texture of the target can be effectively reconstructed. These studies indicate that UAVs provide the advantages of non-contact acquisition, high efficiency, broad coverage, and relatively low cost, thereby establishing a reliable data foundation for the 3D reconstruction of complex targets.

Based on these findings, UAV oblique photogrammetry was investigated in this study for the digital modeling of historic-building exteriors. First, the fundamental principles of UAV-based oblique-photogrammetric reconstruction and point cloud processing are introduced. The Jiuyun Fangding monument is then used as the experimental object, and a multi-altitude, multi-angle UAV image-acquisition scheme is designed. The 3D model is reconstructed through feature matching, camera-pose estimation, dense point cloud generation, point cloud registration, and texture mapping. Finally, the accuracy of the model is validated and analyzed using field-measured dimensions. An integrated UAV-based 3D reconstruction workflow suitable for complex historic-building exteriors is thereby established. Measurable and traceable technical support for the digital documentation, condition assessment, conservation, and restoration of historic buildings can be provided by the results.

2. Theoretical foundations of UAV 3D reconstruction and point cloud processing

2.1. Principles of 3D reconstruction using UAV oblique photography

The 3D reconstruction of UAV oblique photography recovers the 3D information of the target scene through the geometric constraints of multi-view imagery. This reconstruction involves the transformation among four coordinate systems: the world coordinate system (X_w, Y_w, Z_w), the camera coordinate system $(X_c,\ Y_c,\ Z_c)$, the image coordinate system (x, y), and the pixel coordinate system $(u,\ v)$, as illustrated in Figure 1.

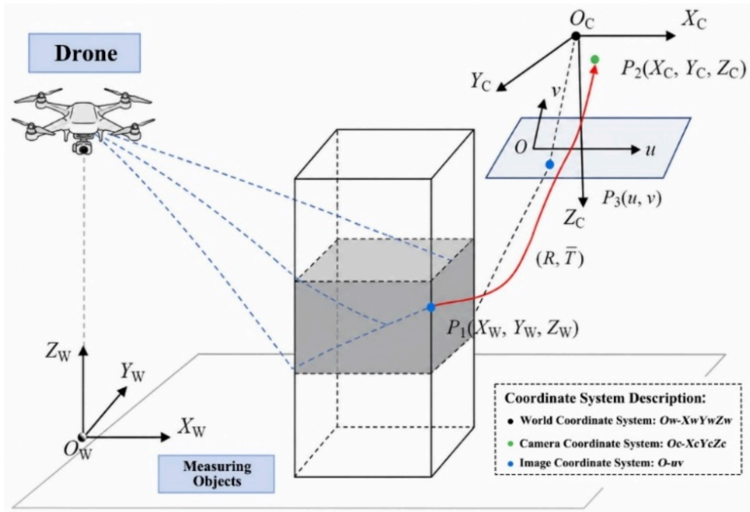

Figure 1. Relationships among coordinate systems in stereo vision

The transformation from the world coordinate system to the camera coordinate system is realized by the rotation matrix R and the translation vector T:

$$\begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix} = \begin{bmatrix} R & T \\ \vec{0} & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \tag{1}$$

where X_w, Y_w, Z_w represent the 3D coordinates of a spatial point in the world coordinate system; X_c, Y_c, Z_c represent the 3D coordinates of the same point after transformation into the camera coordinate system; R is a 3×3 rotation matrix determined by the camera extrinsic parameters, characterizing the rotational attitude of the camera coordinate system relative to the world coordinate system; and T is a 3×1 translation vector (offset matrix) determined by the camera extrinsic parameters, characterizing the translational offset between the origins of the two coordinate systems.

According to the principle of central perspective projection, a 3D point in the camera coordinate system is projected onto the 2D physical image plane. This transformation is primarily governed by the focal length of the camera:

$$Z_c \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} \tag{2}$$

where x and y are the 2D coordinates of the projected point in the image coordinate system, and f is the physical focal length of the camera.

For digital processing, the physical positions (in mm) of the 2D image need to be discretized and mapped to corresponding pixel positions (in pixels). The relationship is expressed as:

$$Z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = Z_c \begin{bmatrix} \frac{1}{dx} & 0 & u_0 \\ 0 & \frac{1}{dy} & \nu_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{3}$$

where u and v are the 2D coordinates of the image point in the pixel coordinate system; dx and dy are the physical dimensions of a single pixel on the camera sensor along the x-axis and y-axis, respectively; u_0 and v_0 represent the offsets from the center of the image coordinate system to the center of the pixel coordinate system.

From Equation (1)-(3), a complete mathematical mapping model from the world coordinate system to the pixel coordinate system can be established:

$$Z_c \begin{bmatrix} u \\ \nu \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{dx} & 0 & u_0 \\ 0 & \frac{1}{dy} & \nu_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ \vec{0} & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \tag{4}$$

Feature extraction is a critical step for achieving matching of corresponding points across multi-view images. To overcome variations in scale, rotation, and illumination among UAV-acquired images, algorithms such as the Scale-Invariant Feature Transform (SIFT) are typically employed to construct a Gaussian scale space.

Multi-scale convolution analysis of the original image is performed using a Gaussian kernel function to construct the Gaussian scale space of the image:

$$L(x_i, y_i, \sigma_i) = G(x_i, y_i, \sigma_i)^* I(x_i, y_i) \tag{5}$$

$$G(x_i, y_i, \sigma_i) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(x-x_i)+(y-y_i)}{2{\sigma_i}^2}\right) \tag{6}$$

where $L(x_i,\ y_i,\ \sigma_i)$ is the i-th layer image in the image space; $G(x_i,\ y_i,\ \sigma_i)$ is the 2D Gaussian blur function; $I(x_i,\ y_i)$ is the pixel value of the original image at position $(x_i,\ y_i)$, σ is the scale-space influence factor; and the operator $*$ denotes the sum of products after convolution of the pixel value with the Gaussian kernel.

A Difference-of-Gaussian (DoG) pyramid is constructed for extreme point detection:

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma))^* I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma) \tag{7}$$

where $D(x, y, \sigma)$ is the generated DoG image; k is the scale ratio factor between adjacent Gaussian scale spaces; and $L(x, y, k\sigma)$ and $L(x,\ y,\ \sigma)$ are Gaussian images at adjacent scales.

To obtain more accurate feature point coordinates, quadratic subpixel interpolation fitting of the scale function in 3D space is performed to eliminate low-contrast and unstable edge response points:

$$F(X) = F(X_0) + \frac{\partial F^T}{\partial X} X + \frac{X^T}{2} \frac{\partial^2 F}{\partial X^2} \tag{8}$$

$$\overline{X} = -\frac{\partial^2 F^{-1}(X)}{\partial X^2} \frac{\partial F(X)}{\partial X} \tag{9}$$

where $F(X)$ is the Taylor expansion of the DoG function at the candidate point; X is the offset vector of the sampling point relative to the central extreme point; and $\overline{X}$ is the precise offset coordinate of the final feature point determined by interpolation fitting.

In the feature point matching stage, to ensure the stability of feature points under translation or rotation, the gradient magnitude $m(x, y)$ and orientation $\theta(x, y)$ of the neighborhood pixels are computed using image gradient methods:

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \tag{10}$$

$$\theta(x, y) = \arctan \frac{(L(x+1,y)-L(x-1,y)}{(L(x,y+1)-L(x,y-1)} \tag{11}$$

After obtaining matched feature points, the camera poses parameters and the true spatial coordinates of the feature points are computed using Structure-from-Motion (SfM). In this process, the image point information in each image is defined as observations of collinearity equations for central projection. The image point coordinates (x, y), camera principal point (x_0, y_0), focal length f, object point coordinates (X, Y, Z), and camera projection center (X_S, Y_S, Z_S) satisfy the collinearity equations:

$$\begin{cases} x - x_0 = -f\frac{a_1(X-X_S)+b_1(Y-Y_S)+c_1(Z-Z_S)}{a_3(X-X_S)+b_3(Y-Y_S)+c_3(Z-Z_S)} \\ y - y_0 = -f\frac{a_2(X-X_S)+b_2(Y-Y_S)+c_2(Z-Z_S)}{a_3(X-X_S)+b_3(Y-Y_S)+c_3(Z-Z_S)} \end{cases} \tag{12}$$

Based on these collinearity equations, bundle block adjustment is performed to integrate the elementary adjustment units. Building upon the sparse point cloud and camera pose parameters generated by SfM, the Multi-View Stereo (MVS) algorithm is further applied for pixel-level dense matching, ultimately generating a high-density point cloud capable of representing the surface details of the ancient building.

Reprojection error is a core metric for evaluating and optimizing the accuracy of 3D reconstruction. It represents the distance between the 2D image plane projection of a 3D spatial point, computed using the estimated camera pose and intrinsic parameters, and the actual extracted image feature point position.

To minimize the reprojection error, the collinearity equations are linearized when the interior orientation elements are known, and the following error equation is constructed:

$$\begin{cases} v_x = & \frac{\partial x}{\partial b}db + \frac{\partial x}{\partial a}da + \frac{\partial x}{\partial c}dc + \frac{\partial x}{\partial x_s}dx_s + \frac{\partial x}{\partial y_s}dy_s + \frac{\partial x}{\partial z_s}dz_s \\ & \quad + \frac{\partial x}{\partial X}dX + \frac{\partial x}{\partial Y}dY + \frac{\partial x}{\partial Z}dZ - I_x, \\ v_y = & \frac{\partial y}{\partial b}db + \frac{\partial y}{\partial a}da + \frac{\partial y}{\partial c}dc + \frac{\partial y}{\partial x_s}dx_s + \frac{\partial y}{\partial y_s}dy_s + \frac{\partial y}{\partial z_s}dz_s \\ & \quad + \frac{\partial y}{\partial X}dX + \frac{\partial y}{\partial Y}dY + \frac{\partial y}{\partial Z}dZ - I_y. \end{cases} \tag{13}$$

where v_x and v_y are the reprojection error residuals in the x and y directions, respectively; and I_x and I_y are the coordinate offsets computed from approximate values. The error equation is solved using bundle adjustment and the least squares method. Camera parameters and 3D point coordinates are iteratively optimized until the reprojection error satisfies the preset threshold. This process generates a high-precision sparse point cloud model.

2.2. Principles of point cloud data processing

Point cloud registration aims to determine the optimal rigid transformation matrix that aligns point cloud data acquired from different viewpoints or different batches into a unified spatial coordinate system. To balance computational efficiency and registration accuracy, a combined strategy of "coarse registration + fine registration" is typically adopted.

Coarse registration provides a favorable initial pose for subsequent fine registration. The Random Sample Consensus (RANSAC) algorithm is employed to extract correct matching correspondences from point clouds containing noise. The required minimum number of iterations k is computed using the following model:

$$k = \frac{\log(1-p)}{\log(1-w^n)} \tag{14}$$

where k is the number of iterations required by the algorithm; p is the confidence probability that a randomly selected point from the dataset during the iteration is an inlier; w is the probability that a point in a randomly selected matching point pair from the dataset is an inlier; N is the number of points randomly sampled each time; and $1-w^n$ represents the probability that at least one of the n points is not a correct matching point pair.

To further optimize the search efficiency for global correspondences, four coplanar points (a,b,c,d) are selected within the same plane, with the requirement that three of the points are non-collinear. Lines ab and cd are connected and intersect at point e, as shown in Figure 2. Using these four points as a point base, geometric constraints are constructed based on the invariance of cross-ratios:

$$r_1=\frac{|a-e|}{|a-b|} \tag{15}$$

$$r_2=\frac{|c-e|}{|c-d|} \tag{16}$$

where r_1 and r_2 are the constructed coplanar four-point cross-ratio constraints.

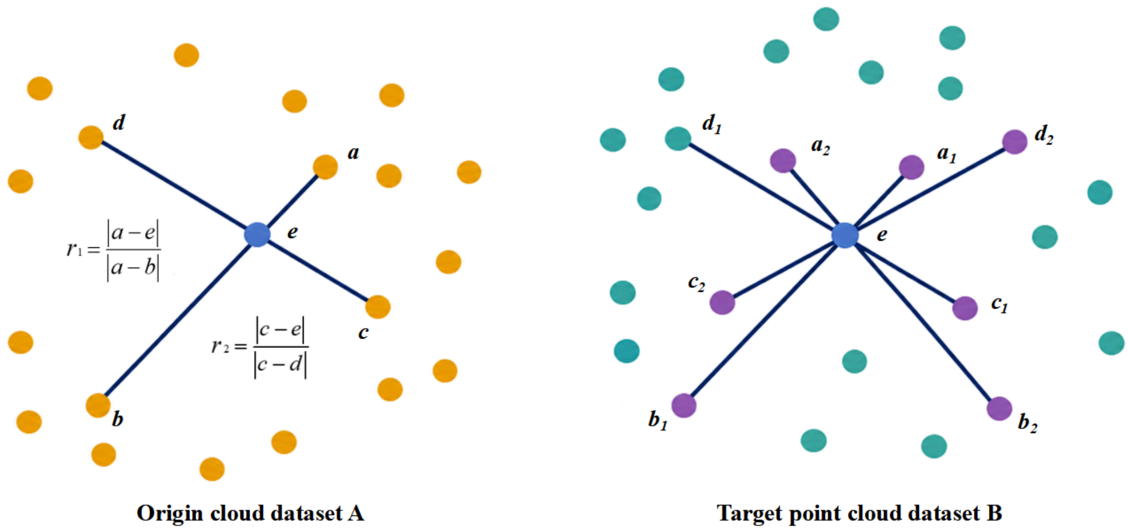

Figure 2. Point cloud dataset

By searching for point bases in the source point cloud and target point cloud that satisfy the matching conditions of ratios r_1 and r_2, the optimal rigid transformation matrix for coarse registration can be rapidly solved using the least squares method.

The initial alignment result provided by coarse registration is insufficient for high-precision measurement requirements; therefore, the Iterative Closest Point (ICP) algorithm is further applied for fine registration. The traditional ICP algorithm optimizes the pose by minimizing the point-to-point Euclidean distance, with the error loss function expressed as:

$$E(R,T)=\frac{1}{k}\sum_{i=1}^{k}(\| q_i-(Rp_i+T)\|^2) \tag{17}$$

where $E(R,T)$ is the registration error loss function; k is the total number of matching point pairs involved in the error calculation; p_i and q_i are the spatial coordinates of corresponding points in the source point cloud and target point cloud, respectively; and R and T are the rotation matrix and translation matrix to be solved.

3. UAV image acquisition and 3D reconstruction experiment for a historic structure

3.1. Experimental object

The Jiuyun Fangding monument, a landmark cultural structure, was selected as the experimental object, as shown in Figure 3. The structure was designed after an ancient Chinese rectangular bronze *ding* and embodies traditional cultural symbolism and classical architectural aesthetics. Its tall, free-standing main body and open surroundings make it well suited to UAV oblique-photogrammetric acquisition and complex-surface reconstruction.

The Jiuyun Fangding monument is a representative complex façade structure with variable cross-sections. Its morphology has two principal characteristics. First, the geometry is complex and incorporates multiple overhanging eaves, columns, and recessed openwork regions, which can produce substantial visual occlusion. Second, the surface texture is highly detailed, with elaborate classical meander patterns and reliefs carved into the main body. These features impose demanding requirements on point cloud matching accuracy and texture-mapping quality. Conventional orthographic or single-circuit flight paths may therefore result in local data loss, making a morphology-based, multi-view, multi-altitude acquisition scheme necessary.

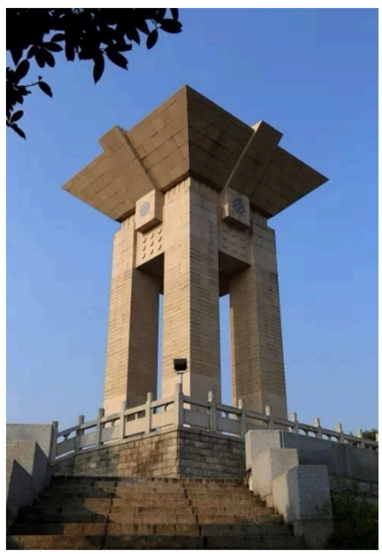

Figure 3. Field photograph of the Jiuyun Fangding monument

3.2. Experimental equipment

Images were acquired using a DJI Mavic 3 Enterprise professional surveying UAV, as shown in Figure 4. Its specifications are presented in Table 1. The aircraft is equipped with a high-accuracy Real-Time Kinematic (RTK) positioning module and provides strong wind resistance and stable hovering performance. Its high-resolution camera incorporates a 4/3 Complementary Metal-Oxide-Semiconductor (CMOS) sensor with 20 effective megapixels, a 24-mm equivalent focal length, and an adjustable gimbal-pitch range of -90° to 35°. These characteristics allow fine relief details on the building surface to be captured clearly.

Figure 4. DJI Mavic 3 Enterprise UAV used in the experiment

Table 1. Specifications of the DJI Mavic 3 Enterprise UAV

Device	Parameter	Value
UAV	Weight	1,050g
	Maximum flight duration	45 min under windless test conditions
	Unfolded dimensions	347.5×283×107.7mm
	Maximum hovering duration	38 min
Camera	Wide-angle camera sensor	4/3 CMOS,20 effective megapixels
	Maximum image dimensions	5,280×3,956 pixels
	Wide-angle lens	24mm equivalent focal length
	Gimbal-pitch range	-90°~35°

To ensure the absolute spatial accuracy of the 3D model and establish a reference for subsequent evaluation, a Global Navigation Satellite System–Real-Time Kinematic (GNSS–RTK) receiver was used to establish ground control points and acquire their 3D coordinates. A high-accuracy total station was also used to measure representative dimensions of the Jiuyun Fangding monument in the field, including the relevant column dimensions, column spacings, and stair-step height. These measurements were subsequently compared with the corresponding dimensions extracted from the reconstructed model.

3.3. UAV image-acquisition scheme

As illustrated in Figure 5, an optimized flight-planning strategy combining layered fixed-position circumferential flights with oblique photography was adopted. The center of the structure was used as the target anchor, and the UAV was flown along multiple circumferential paths at different altitudes and appropriate stand-off distances from the target surface. Larger downward viewing angles were used to acquire the horizontal surfaces and top-view textures of the roof and base. Nearly horizontal viewing angles were used during circumferential imaging of the main façades, thereby producing panoramic viewpoints with minimal blind areas.

To satisfy the feature-extraction requirements of the SfM algorithm and prevent holes from appearing in the reconstructed openwork or relief regions, the forward and side overlaps were set to 80% and 75%, respectively. Through frequent shutter triggering and appropriate adjustment of the spacing between flight paths, each point on the building surface was clearly imaged in at least four to five photographs.

Field flights were conducted under relatively uniform illumination, without strong backlighting, and during periods of low wind speed. Ground control points were distributed uniformly around the target area. The UAV was then operated along the predetermined routes, and supplementary multi-view images were acquired at fixed positions where necessary. After blurred and redundant images had been removed, the remaining valid images were retained for reconstruction.

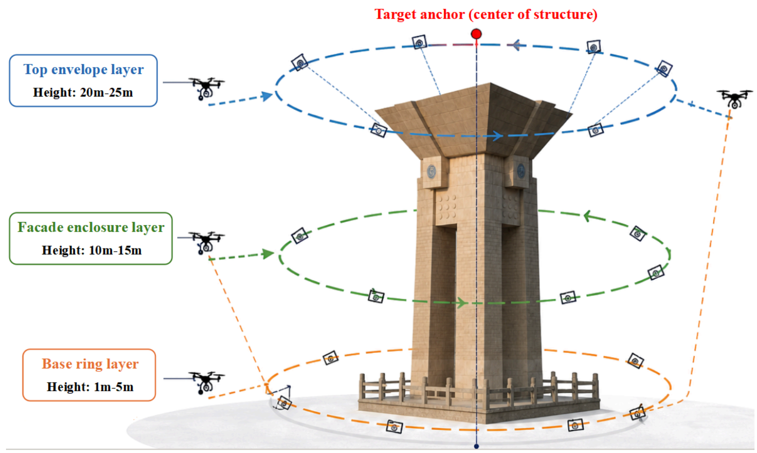

Figure 5. Schematic of the layered circumferential UAV flight plan

As shown in Figure 6, the UAV was used to acquire on-site images of the target structure. A spiral circumferential flight path was established around the exterior of the building, and images were continuously captured from different altitudes and angles as the UAV moved upward. A multi-view dataset with high image overlap was thereby obtained. This acquisition strategy provided relatively complete coverage of the façades, roof, and individual structural components, establishing the data foundation required for subsequent feature matching, dense point cloud generation, and 3D model reconstruction.

Figure 6. On-site UAV image acquisition

4. Accuracy validation and analysis of the 3D model

4.1. 3D model reconstruction results

After UAV image acquisition had been completed, the quality of the raw data was examined. Images affected by flight vibration, abnormal exposure, or unsuitable viewing angles were removed. The coordinate information associated with the retained images was then standardized. Local feature points were then extracted from the images using SIFT. Stable feature detection could therefore be maintained despite moderate variations in image scale, rotation, and illumination. Correspondences among homologous points in images captured from different viewpoints were established by comparing the feature descriptors. Incorrect matches were subsequently removed using RANSAC, providing reliable correspondence data for camera-pose estimation and 3D point recovery.

After the homologous feature points had been obtained, the camera position and orientation associated with each image were calculated using SfM. The 3D coordinates of the feature points were recovered according to central-projection geometry, producing an initial sparse point cloud of the Jiuyun Fangding monument. Bundle adjustment was then performed by minimizing the reprojection error. The camera intrinsic and extrinsic parameters and spatial point coordinates were jointly optimized. This reduced the influence of image-matching errors on the geometric accuracy of the model.

After the sparse point cloud and camera poses had been determined, pixel-level dense matching was performed using MVS. A dense point cloud representing the main body, base, load-bearing columns, and decorative textures was thereby generated. Potential local displacements between point clouds obtained from different viewpoints or acquisition batches were corrected through coarse and fine registration. Outlier removal, point cloud fusion, hole filling, surface-mesh construction, and texture mapping were subsequently performed to obtain the complete 3D model.

The model before and after processing is compared in Figure 7. The initial model represented the basic outline of the structure; however, point dispersion, surface gaps, and local overlap remained in several areas because of occlusion, repetitive textures, and differences among the camera viewpoints. After dense matching, point cloud registration, and surface reconstruction had been performed, the number of dispersed points was substantially reduced. The spatial relationships among the main body, load-bearing columns, base, and stairs became more continuous. In addition, local decorative textures and edge contours were represented more completely. The final model not only provided an intuitive visualization of the overall form of the Jiuyun Fangding monument but also supported dimensional measurement and spatial analysis, thereby providing a basis for model-accuracy validation.

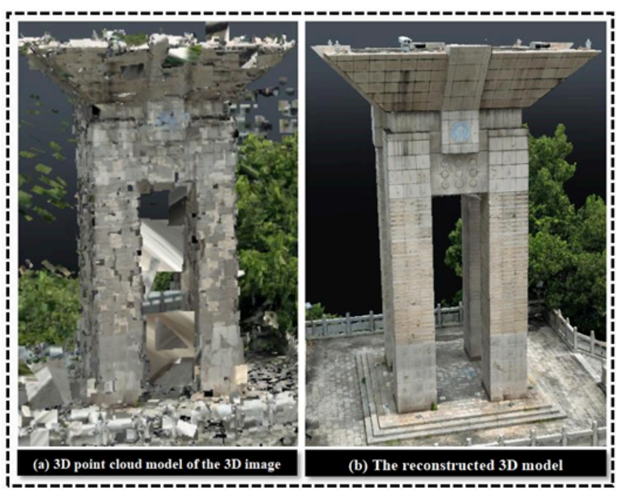

Figure 7. Comparison of the Jiuyun Fangding model before and after reconstruction processing

4.2. Model accuracy validation

To verify the dimensional reliability of the UAV-based 3D reconstruction results, four representative geometric features on the Jiuyun Fangding that were amenable to on-site re-measurement were selected as check objects: the width of the load-bearing columns, the minimum spacing between load-bearing columns, the maximum spacing between load-bearing columns, and the height of the stair steps. The dimensions of each check feature were extracted from the 3D model, and the on-site measurement results were used as the benchmark for comparison. The evaluation metrics employed were absolute error and relative error: absolute error reflects the numerical deviation between the model-derived measurement and the measured value, while relative error measures the proportion of this deviation relative to the measured dimension. Figure 8 illustrates the measurement positions of each check feature in the point cloud model and their comparison with the on-site measurement results.

Figure 8. Comparison between field measurements and point cloud measurements of the Jiuyun Fangding monument: (a) load-bearing-column width; (b) minimum spacing between the load-bearing columns; (c) maximum spacing between the load-bearing columns; and (d) stair-step height

The model-derived and field-measured values are summarized in Table 2. The model-derived load-bearing-column width was 1.793 m, whereas the field-measured value was 1.826 m, corresponding to an absolute error of 0.033 m and a relative error of 1.81%. The model-derived minimum spacing between the load-bearing columns was 2.564 m, corresponding to an absolute error of 0.024 m and a relative error of 0.93%. The model-derived maximum spacing between the load-bearing columns was 6.161 m, whereas the field-measured value was 6.200 m, corresponding to an absolute error of 0.039 m and a relative error of 0.63%. The model-derived stair-step height was 0.447 m, whereas the field-measured value was 0.453 m, corresponding to an absolute error of 0.006 m and a relative error of 1.32%. Thus, the absolute errors of the four validation dimensions ranged from 0.006 to 0.039 m, and all relative errors remained below 2%.

Table 2. Comparison between the model-derived and field-measured dimensions

ID	Description	Model-derived value (m)	Field-measured value (m)	Absolute error (m)	Relative error (%)
A	Load-bearing-column width	1.793	1.826	0.033	1.81
B	Minimum spacing between the load-bearing columns	2.564	2.588	0.024	0.93
C	Maximum spacing between the load-bearing columns	6.161	6.200	0.039	0.63
D	Stair-step height	0.447	0.453	0.006	1.32

The relationships among the model-derived values, field-measured values, absolute errors, and relative errors are further illustrated using bar charts in Figure 9. As shown in Figure 9(a) and 9(b), the model-derived measurements of the four validation features were generally close to the corresponding field measurements, indicating satisfactory agreement between the two datasets. Figure 9(c) shows that the maximum spacing between the load-bearing columns had the largest absolute error, at 0.039 m, whereas the stair-step height had the smallest absolute error, at only 0.006 m. As shown in Figure 9(d), the load-bearing-column width had the largest relative error, at 1.81%, whereas the maximum column spacing had the smallest relative error, at 0.63%.

These results indicate that the absolute error was influenced to some extent by the actual size of the measured component. A larger component may exhibit a greater absolute deviation without necessarily exhibiting a greater relative error. The accuracy of a reconstructed model should therefore be evaluated using both absolute and relative errors.

Model errors were primarily attributable to image acquisition and data processing. During image acquisition, minor variations in the UAV flight attitude, camera-imaging errors, illumination differences, and local occlusion may have affected the extraction quality of homologous feature points. In addition, repetitive decorative textures, recessed regions, and component edges on the Jiuyun Fangding monument may have caused unstable local matching. During data processing, errors may have been introduced through point cloud filtering, registration of point clouds from different viewpoints, surface fitting, and the manual selection of measurement endpoints. Furthermore, small-scale components such as stair steps were represented by relatively few valid point cloud observations. Consequently, even a small coordinate deviation could produce a comparatively large relative error.

Despite these potential sources of error, the relative errors of all four validation dimensions were below 2%, indicating that the reconstructed model provided stable geometric consistency. Compared with manual point-by-point measurement, UAV-based 3D reconstruction offers non-contact operation, broad coverage, high data-acquisition efficiency, and repeatable measurement. The method is particularly suitable for acquiring data from elevated areas, roofs, and regions of historic buildings that are difficult for personnel to access. Both the overall architectural form and local surface textures can be recorded during a single acquisition campaign.

Thus, the survey results can be extended from a small number of discrete measurements to preservable, traceable, and repeatedly analyzable 3D spatial data. These characteristics demonstrate the advantages of the proposed method for digital documentation, exterior visualization, dimensional verification, and conservation-condition assessment.

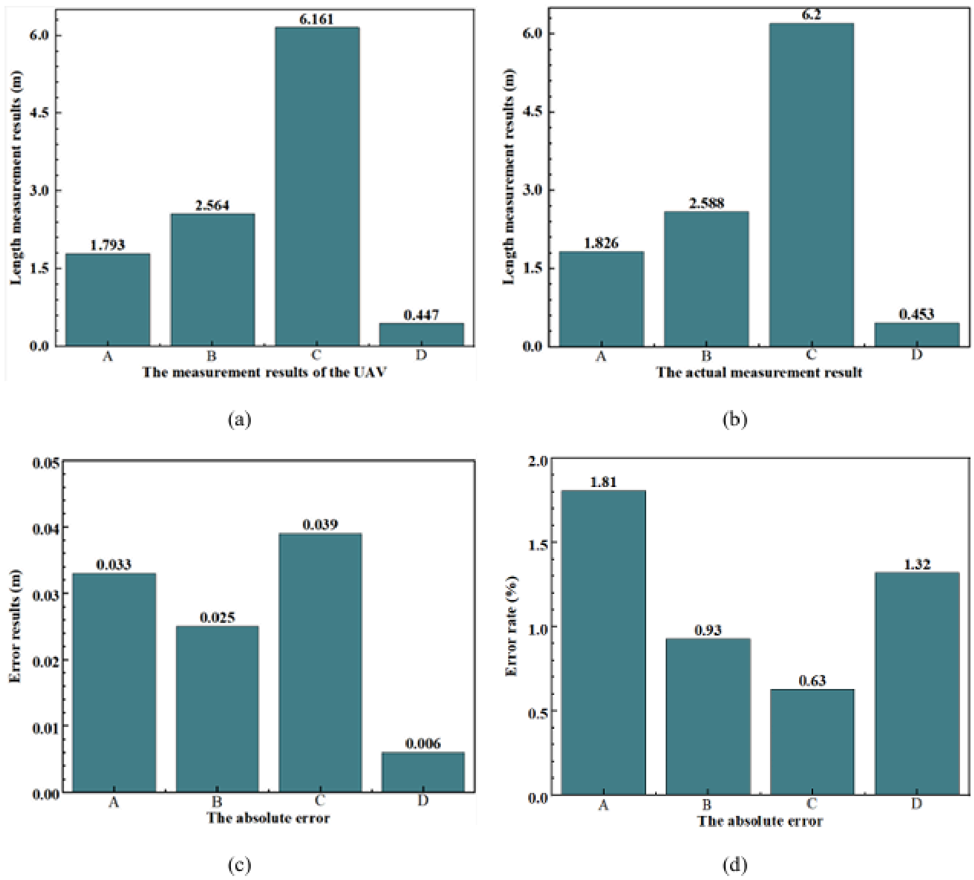

Figure 9. Bar chart comparison of the measured dimensions: (a) model-derived values; (b) field-measured values; (c) absolute errors; and (d) relative errors

5. Conclusions

The Jiuyun Fangding monument was selected as the study object, and an integrated technical workflow incorporating UAV image acquisition, 3D point cloud generation, model reconstruction, and accuracy validation was established. The principal conclusions are as follows.

(1) Images were acquired using a DJI Mavic 3 Enterprise UAV equipped with a 20-megapixel camera. A layered circumferential acquisition strategy incorporating multiple altitudes and viewing angles was adopted. The forward and side overlaps were set to 80% and 75%, respectively, allowing the principal areas of the building surface to be imaged repeatedly in four to five photographs. The façades, roof, base, and local structural components were covered relatively comprehensively, and data loss caused by single-view imaging was reduced.

(2) The 3D model of the Jiuyun Fangding monument was reconstructed through image-feature extraction, homologous-point matching, SfM, MVS matching, and coarse-to-fine point cloud registration. After outlier

removal, point cloud fusion, hole filling, and texture mapping had been performed, the point dispersion, local overlap, and surface gaps observed in the initial model were substantially reduced. The final model clearly represented the main body, load-bearing columns, base, stairs, and other principal components and supported the measurement and analysis of four representative dimensions.

(3) The accuracy-validation results showed that the absolute errors of the four dimensions ranged from 0.006 to 0.039 m, while the relative errors ranged from 0.63% to 1.81%. The relative error of every validation item was below 2%. The maximum spacing between the load-bearing columns exhibited the smallest relative error, at 0.63%, whereas the load-bearing-column width exhibited the largest relative error, at 1.81%. The model-derived dimensions agreed well with the field measurements, indicating that the model can satisfy the requirements of digital exterior documentation, 3D visualization, and general dimensional verification for historic buildings.

The present study was limited to a single structure and four geometric validation dimensions. Future investigations should include historic buildings with different structural forms and surface textures. Reconstruction accuracy for complex components and local details could be further improved by optimizing the distribution of ground control points, acquiring additional close-range images of occluded areas, and introducing 3D laser-scanning data for cross-validation.

References

[1] Wen, X., & Wu, G. (2022). Heterogeneous multi-drone routing problem for parcel delivery. *Transportation Research Part C: Emerging Technologies, 141*, Article 103763. https: //doi.org/10.1016/j.trc.2022.103763

[2] Wu, Y., Low, K. H., Pang, B. Z., & Tan, Q. Y. (2021). Swarm-based 4D path planning for drone operations in urban environments. *IEEE Transactions on Vehicular Technology, 70*(8), 7464-7479. http: //doi.org/10.1109/tvt.2021.3093318.

[3] Ding, G. R., Wu, Q. H., Zhang, L. Y., Lin, Y., Tsiftsis, T. A., & Yao, Y. D. (2018). An amateur drone surveillance system based on the cognitive internet of things. *IEEE Communications Magazine, 56*(1), 29-35. http: //doi.org/10.1109/mcom.2017.1700452.

[4] Tang, W., Zhao, H., Chen, S., Li, A., Yang, W., Cheng, W., Li, Y., & Dong, L. (2025). Measurement of LoRa signal propagation in urban areas utilizing aerial gateway and ground gateway. *Scientific Data, 12*(1), Article 1464. https: //doi.org/10.1038/s41597-025-05802-2

[5] Zhang, J. T., Cheng, Q. F., Chen, X. F., & Luo, X. Y. (2025). CSAP-IoD: A chaotic map-based secure authentication protocol for internet of drones. *IEEE Transactions on Information Forensics and Security, 20*, 8848-8862. http: //doi.org/10.1109/tifs.2025.3599678.

[6] Meng, X., Zhu, Z., Liu, H., Lian, Y., Lu, H., Wang, Y., Shi, R., & Spencer, B. F. (2025). A drone-borne GPR system for lake and river ice thickness monitoring. *Water Resources Research, 61*, Article e2025WR040290. https: //doi.org/10.1029/2025WR040290

[7] Li, J., Ni, X., Hua, M., Wang, Q., Wu, Q., Zhao, X., & Xu, M. (2026). Energy-efficient large-scale low-altitude delivery: Integrating charging station location and UAV route planning by a cascade optimization framework. *Aerospace Science and Technology, 168*, Article 111119. https: //doi.org/10.1016/j.ast.2025.111119

[8] Gu, R., Zhao, Y., & Ren, X. (2026). Integrating wind field analysis in UAV path planning: Enhancing safety and energy efficiency for urban logistics. *Chinese Journal of Aeronautics, 39*(1), Article 103605. https: //doi.org/10.1016/j.cja.2025.103605

[9] Chen, J., Huang, J., Zhang, Z., Yang, J., Wu, Z., & Luo, R. (2026). ATM-Net: A lightweight multimodal fusion network for real-time UAV-based object detection. *Drones, 10*(1), Article 67. https: //doi.org/10.3390/drones10010067

[10] Piatkowska, K., & Darecka, K. (2026). Restoring the past digitally: A 3D visualization study of the Great Mill in Gdansk. *Design Studies, 103*, Article 101376. https: //doi.org/10.1016/j.destud.2025.101376

[11] Balletti, C., & Ballarin, M. (2019). An application of integrated 3D technologies for replicas in cultural heritage. *ISPRS International Journal of Geo-Information, 8*(6), Article 285. https: //doi.org/10.3390/ijgi8060285

[12] Liu, Z. (2021). Extraction and mapping of component information of ancient buildings in Huizhou based on UAV technology. *Ecological Informatics, 66*, Article 101437. https: //doi.org/10.1016/j.ecoinf.2021.101437

[13] Wang, M., Xu, J., Lassiter, H. A., & Li, S. (2025). BIM-driven laser spatial augmented reality for in-situ layout and assembly. *Automation in Construction, 178*, Article 106405. https: //doi.org/10.1016/j.autcon.2025.106405

[14] Lovrinčević, M., Papa, I., Janeš, D., Hodak, L., Pentek, T., & Đuka, A. (2025). New possibilities of field data survey in forest road design. *Sensors, 25*(13), Article 4192. https: //doi.org/10.3390/s25134192

[15] Jia, H., & Rong, W. (2026). Efficient sub-pixel laser centerline extraction via an improved U-Net for structured light measurement. *Measurement, 257*, Article 118807. https: //doi.org/10.1016/j.measurement.2025.118807

[16] Wang, H., Dang, Z., Cui, M., Shi, H., Qu, Y., Ye, H., Zhao, J., & Wu, D. (2025). ADG-YOLO: A lightweight and efficient framework for real-time UAV target detection and ranging. *Drones, 9*(10), 707. https: //doi.org/10.3390/drones9100707

[17] Jia, Z., Wang, B., & Chen, C. (2024). Drone-NeRF: Efficient NeRF based 3D scene reconstruction for large-scale drone survey. *Image and Vision Computing, 143*, Article 104920. https: //doi.org/10.1016/j.imavis.2024.104920

[18] Lu, J., Liu, Y. M., Jiang, C. M., & Wu, W. W. (2025). Truck-drone joint delivery network for rural area: Optimization and implications. *Transport Policy, 163*, 273–284. http: //doi.org/10.1016/j.tranpol.2025.01.016