

A UAV PHOTOGRAPHIC PATH PLANNING METHOD FOR HIGH-QUALITY RECONSTRUCTION OF CULTURAL HERITAGE

Hao Zhou ¹, Yuchen Liu ¹, Zheng Ji ^{1*}

¹ School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China
- (haozhu47, liuyc092, jz07)@whu.edu.cn

KEY WORDS: Cultural Heritage, 3D Reconstruction, UAV Path Planning, Shaped DSM, Greedy Optimization

ABSTRACT:

The image-based reconstruction method can preserve geometric and textural information with relatively high accuracy, making it a suitable method for digitally documenting cultural heritage. However, the quality of the reconstructed model largely depends on the quality of the captured images. Unmanned aerial vehicles (UAVs) equipped with a camera and gimbal offer great convenience for image acquisition in 3D reconstruction. However, ensuring safety, high efficiency, and full coverage is a challenge. To address this, we propose a UAV photographic path planning method for efficient and automatic image acquisition of heritage scenes, based on which high-quality reconstruction is realized. A priori proxy of the scene is obtained in advance and utilized to (1) generate initial viewpoints for subsequent optimization; (2) generate the SDSM for obstacle avoidance, signal analysis, and sight occlusion judgment; and (3) segment to obtain planar regions to sample representative points for measuring the reconstructability of heritage scene and optimizing the viewpoints. Our method enables the planning of regular and safe final paths for the high-quality reconstruction of cultural heritage, outperforming both commercial software and state-of-the-art methods in both real and virtual scenes.

1. INTRODUCTION

In recent years, the demand for documentation of cultural heritage is constantly increasing (Aicardi et al., 2018; Gomes et al., 2014; Murtiyoso and Grussenmeyer, 2017). The image-based reconstruction method can better preserve the geometric and textural information at a low cost and has become an important tool for heritage documentation (El-Hakim et al., 2004; Liu et al., 2022). Now that the image 3D reconstruction algorithms are more mature, the image quality becomes the main factor affecting the reconstruction quality. Inadequate and insufficient coverage can result in mismatches between images and holes in reconstructed models (Furukawa and Hernández, 2015; Schönberger et al., 2016), while excessively redundant images would increase the time and calculation cost during image acquisition and reconstruction processes, and even lead to poor reconstruction quality (Zhang et al., 2021). Considering the complexity of the cultural heritage scene, how to capture images is an essential issue in its reconstruction.

Currently, commonly used image acquisition methods for large scenes include conventional vertical aerial photography and oblique photography (Dahlke et al., 2015; Nesbit and Hugenholtz, 2019). For flight safety, these two methods usually carry out a regular path in a grid, zigzag or circle path beyond a certain distance above the scene. They are mostly carried out with a 2D horizontal flight or 2.5D terrain-following flight (Liu et al., 2022; Pepe et al., 2018). Although oblique photography is performed through a tilt camera module with a multi-directional lens, the closer it gets to the bottom, the more texture is missing and distorted in the tilted image. In addition, the occlusion and loss of details caused by remote photography are not addressed (Toschi et al., 2017). High-quality 3D reconstruction requires complete coverage of all details of the heritage scene. 3D close-up photography methods that can completely obtain information from multiple angles are undoubtedly superior (Schönberger et al., 2016; Seitz et al., 2006). Currently, the multi-rotor UAV

equipped with RTK (Real Time Kinematic) module and gimbal provides the hardware implementation basis for close-up photography and automatic image collection (Koch et al., 2019; Li et al., 2017; Nex and Remondino, 2014).

However, most of the 3D flight paths of UAVs are performed under manual control or some predefined flight modes in practical operations. Manual control requires UAV operators to capture images with full coverage and safety guaranteed, making it a complex and challenging task (Zhang et al., 2021). In this situation, a greater number of images is prone to be captured and so cause redundancy and long-time consumption. Thus, the path planning of UAV automatic photography in 3D space is vital for realizing high-quality 3D reconstruction (Hepp et al., 2019; Roberts et al., 2017). It is necessary to plan the photographic position and orientation according to the requirements of 3D reconstruction (Smith et al., 2018; Zhou et al., 2020).

In this paper, we propose a UAV photographic path planning method that can efficiently and automatically complete the image acquisition with safety guaranteed, based on which high-quality and efficient reconstruction of cultural heritage is realized. To verify the effectiveness of our method, we compared it with the state-of-the-art in virtual scenes and commercial software in real scenes. Experiments demonstrate the remarkable performance of our method, which provides a reliable and efficient solution for the high-quality reconstruction of cultural heritage from the aspect of data acquisition.

2. METHODS

In contrast to real-time and online planning methods that do not require a proxy, the photographic planning method discussed in this paper requires a priori proxy of the heritage scene. The overall process from photographic planning, image acquisition to the final 3D reconstruction is complex. We summarize the processes as shown in Eq. (1). Firstly, according to the

* Corresponding author

requirement of model resolution, the appropriate image acquisition parameters $paras$ need to be set, including camera parameters and the ground sample distance (GSD), which corresponds to photographic distance. Then, based on the proxy P of the heritage scene S , the viewpoint set is planned and connected into a collision-free path for image acquisition, from which the model M is reconstructed.

We focus on path planning to improve image acquisition efficiency, ensuring flight safety and high-quality reconstruction. The workflow of our method is shown in Fig. 1.

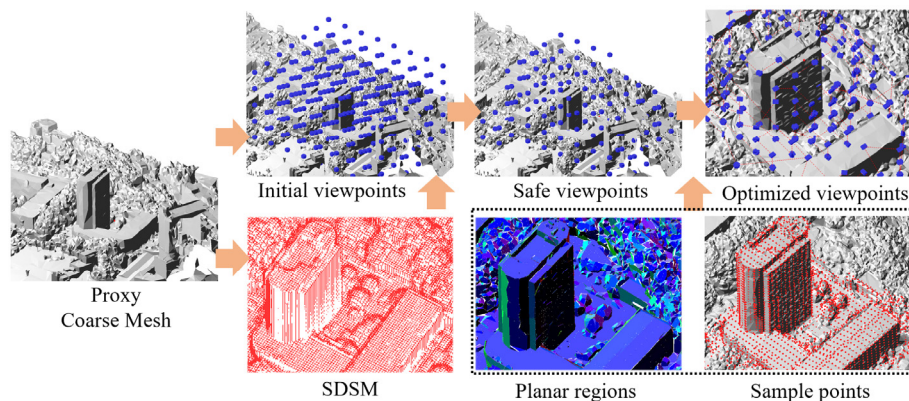


Figure 1. The workflow of UAV path planning method for high-quality reconstruction.

2.1 Viewpoint Generation for Flight Safety and Continuity

Priori Geometric Proxy Similar to the method in (Smith et al., 2018; Zhang et al., 2021; Zhou et al., 2020), our method requires an initial proxy that would serve as the data basis for the entire path planning process, including sample point and viewpoint generation, obstacle avoidance, and occlusion judgment. A relatively accurate absolute position is required for the initial proxy to ensure flight safety and one-to-one correspondence between the planned viewpoint and the actual photographic position. Common proxies include point cloud (Yan et al., 2021), DEM (Smith et al., 2018), 2.5D coarse model (Zhang et al., 2021; Zhou et al., 2020), BIM, 3D mesh obtained by oblique photography (Li et al., 2023), etc. The 3D coarse mesh is usually adopted as the initial proxy, as shown in Fig. 1, which can be reconstructed by capturing images over the scene following a regular path.

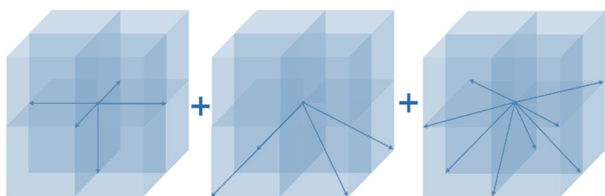


Figure 2. The different orientations generated in every voxel.

Initial Viewpoints Generation We generate regular voxels in the enclosing box of the proxy with the interval $d = 1/3 * W$, where W is the length of the image width projection on the ground. The interval corresponds to the 66% side overlap ratio in aerial photogrammetry, which is considered a relatively high overlap ratio to recover the terrain information and obtain DEM and DOM. In each voxel, 17 viewpoints (a $1*5$ vector representing the camera position and orientation when capturing

$$M = Recon(Capture(Plan(P(S), paras))) \quad (1)$$

The priori proxy, that is the coarse mesh of the scene, is utilized to (1) generate initial viewpoints which are regularly arranged; (2) generate the SDSM for obstacle avoidance and signal analysis of the initial viewpoint set to delete unsafe viewpoints; and (3) segment planar regions to sample representative points for measuring the reconstructability of heritage scene. Based on the reconstructability and its distribution of sample points, the viewpoints with high contribution to the reconstruction are selected to the optimized set during iterations.

an image) are generated for capturing images by taking the center of the voxel and 17 angles facing different directions to form the initial set of viewpoints, as shown in Fig. 2. Considering the complexity of subsequent optimization, certain fixed angles can be removed, and 9 or 13 orientations are used.

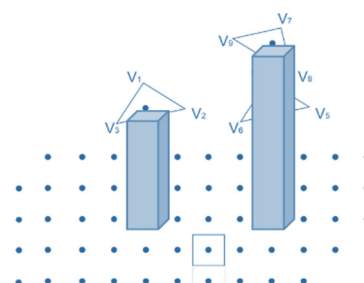


Figure 3. The generation process of SDSM from the proxy.

Shaped DSM Both the generated viewpoints and the final paths connecting the viewpoints need to take into account obstacle avoidance and occlusion. In addition, when judging the visibility of a viewpoint to a target point in the scene, the occlusion of the sight also needs to be considered. Based on the priori proxy of the scene, obstacle avoidance and occlusion can be determined with its geometric structure. The proxy represents the surface to be reconstructed, which is also considered the obstacle. Theoretically, to judge whether the sight and path are occluded or not, it is only necessary to determine if they intersect with the triangular faces of the proxy. However, although the priori model is relatively coarse, the number of faces is still large, especially in large scenes, and the repeated calculation of ray intersection with triangular faces can be computationally intensive. Subsequent signal analysis is also difficult to achieve through intersection judgments. Thus, we convert the proxy to SDSM (Shaped Digital Surface Model) by ray intersection with the

proxy for quick judgment, as shown in Fig. 3. The maximum elevation of every grid of $1 \times 1\text{m}$ on SDSM is retained. In addition, we filter the elevation of SDSM by the maximum value according to the ideal signal mechanism with the empirical value 3:1. That is when the elevation of a certain grid of the SDSM is $3D\text{ m}$, viewpoints within its $n\text{ m}$ range should be higher than $3D - 3 \times n\text{ m}$. The elevation of SDSM represents the minimum flight height of the corresponding range. Based on the SDSM, obstacle avoidance and RTK signal analysis are considered to ensure flight safety and continuity.

A viewpoint is outside the obstacles and in areas with good RTK signal if its elevation is higher than the elevation of SDSM with the same coordinates on the XOY plane, as shown in Fig. 4. For sight and path occlusion judgment, sampling is performed on them to generate points. If the elevation of every point is higher than SDSM, it's believed there is no occlusion, as shown in Fig. 4.

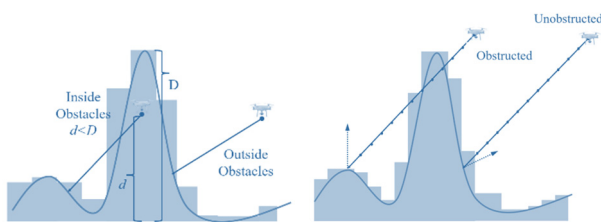


Figure 4. Obstacle avoidance and occlusion judgment

2.2 Viewpoint Optimization for Efficiency Image Acquisition and Reconstruction

Reconstruction Heuristics For viewpoint optimization, the information gained from each viewpoint is required to be measured, as well as the reconstructability of the scene. The common solution is to evaluate the reconstructable value of dense sample points sampling on the proxy based on reconstruction heuristics to represent the reconstructability of the whole scene. The heuristics brought by (Smith et al., 2018) is widely used, in which shooting distance, observation angle, parallax angle and multi-view observation are all taken into account. $H(s, V)$, which is referred to as H in after, represents the reconstructable value of sample point s from the viewpoint set V . According to (Liu et al., 2021; Smith et al., 2018; Zhou et al., 2020), when H of the

sample point exceeds the minimum threshold H_{Tmin} , which is set to 1.3, the sample point is considered to meet the reconstructability. And if H exceeds the maximum threshold H_{Tmax} , which is set to 5.0, redundancy exists in the viewpoints visible to it.

Representative Sampling Many methods such as (Hepp et al., 2019; Roberts et al., 2017; Smith et al., 2018) have attempted to propose more reliable reconstruction heuristics, establishing a strong correlation between the reconstructable value and reconstruction accuracy of sample points. However, we focus on generating typical sample points that are more representative of the scene reconstructability.

Planar segmentation enables the clustering of triangular faces on the priori proxy that are adjacent and approximately oriented in the same direction, and the edges of triangular faces with large variations in normal vectors can also be extracted. Therefore, the proxy is first partitioned into planar regions by the segmentation method of (Bouzas et al., 2020), and geometric primitives including polygons (representing the planar regions) and lines (representing the edges between planar regions) are then extracted. Every triangular face on the proxy belongs to a particular primitive. Even when the heritage scene is complex, the extracted primitives can still reflect the geometry structure of the scene well. We then uniformly sample points on the geometric primitives. The normal vector of the sample point in polygon primitives is consistent with segmented planar regions, while the normal vector of the sample point in line primitives is the average of its two adjacent regions.

Since the normal vectors within the primitives are consistent, the sample points are more representative of the reconstructability of a certain range of their surrounding neighborhood. Thus, a larger interval can be set to generate the sample points, achieving a smaller number of typical sample points to measure the reconstructability of the whole scene. Note that every triangular face on the proxy is clustered into specific planar regions, i.e., each triangular face has a corresponding sample point to measure its reconstructability, and edges with large variations in normal vectors are also sampled. The resulting sample points can cover all the details in the scene to be reconstructed, including edges, corners, fragmentary details, etc., without being particularly dense.

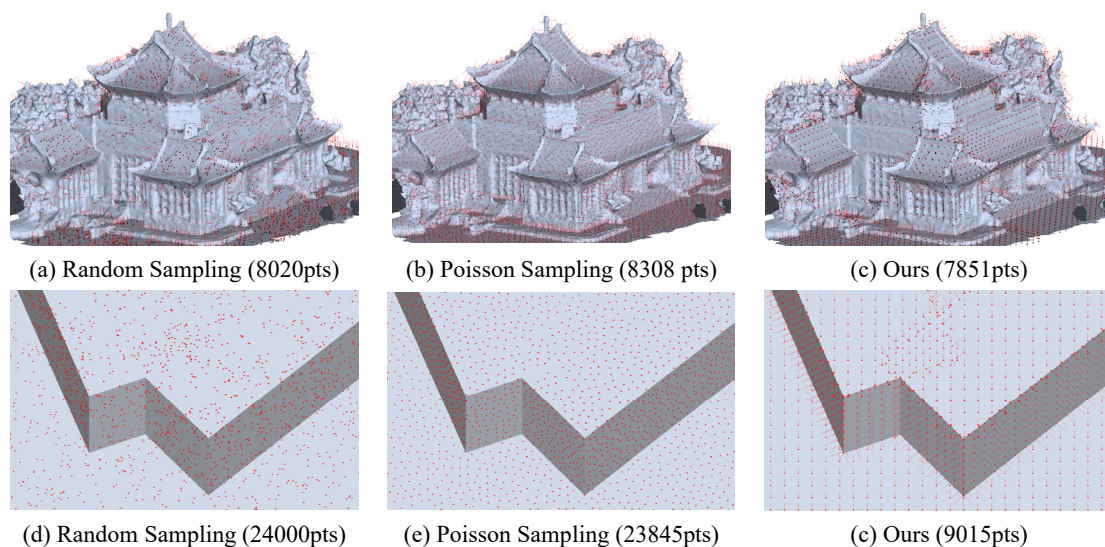


Figure 5. Sample points generated by different sampling methods in 3D proxy (top row) and 2.5D proxy (bottom row)

Fig. 5 shows the distribution of sample points generated by our representative sampling method and the other two methods. Our sample points are evenly distributed in the planar region, with points sampling on discontinuities, which is presented especially with a 2.5D coarse proxy. Compared with the other two sampling methods, a more representative evaluation which includes the reconstructability measurement of every fragment detail is realized with fewer sample points. The validity of our sampling method is verified experimentally subsequently.

Visibility Judgement For fast optimization, we calculate and save the visibility between viewpoints and sample points in advance. First, it is considered that the sample point is not visible from the viewpoint if the distance between them exceeds two times the photographic distances corresponding to the preset GSD. Second, it is not visible if the shooting angle between the shooting sight and the normal vector of the sample point exceeds 60° which leads to a narrow view. Third, a colinear function from photogrammetry is carried out to determine whether the sample point is photographed by the viewpoint. And the last is the sight occlusion judgment in Fig. 4. The visibility between viewpoints

and sample points is determined after the above four conditions are calculated.

Greedy Optimization To quickly optimize the viewpoints, we then construct an objective optimization function as shown in Eq. (2), and continuously iterate to select the viewpoint that obtains the maximum expected gain of reconstructability to join the final optimized set. S_{v_i} is the sample point set which is visible from the viewpoint v_i . V is the existing optimized viewpoint set during the current iteration, and $|V|$ equals the iteration times. $D_{avr}(v_i, V)$ is the average distance from v_i to the optimized viewpoint set. When $V = \emptyset$, we set $D_{avr}(v_i, \emptyset) = \max(v_i, v_j), v_i, v_j \in U$, that is, the maximum distance between two safe viewpoints. U is the safe viewpoint set. $D_{avr}(v_i, S_{v_i})$ is the average distance from v_i to its visible sample point set S_{v_i} . $|S_{v_i} \cap S_V|$ is the cardinality of the intersection set of the set S_{v_i} and S_V . S_V is the sample point set visible from the optimized viewpoint set. A large cardinality value means more intersections with V . θ_{avr} is the average shooting angle of the viewpoint v_i to S_{v_i} . D_{gsd} is the shooting distance corresponding to the preset GSD.

$$v_i = \arg \max \Delta_{EH}(v_i|V), V \subseteq U$$

$$\Delta_{EH}(v_i|V) = \frac{|S_{v_i} \cap S_V| * \cos(\theta_{avr}) * D_{avr}(v_i, V)}{D_{avr}(v_i, S_{v_i})^2 * |D_{avr}(v_i, S_{v_i}) - D_{gsd}|} * \sum_{s \in ur(S_{v_i})} (H_{Tmax} - H(s, V)) \quad (2)$$

The practical implication of Eq. (2) is to find the best viewpoint $v_i = \arg \max \Delta_{EH}(v_i|V)$, whose sum reconstructable value of the unreconstructed sampling point set $ur(S_{v_i})$ is the lowest, i.e. $\sum_{s \in ur(S_{v_i})} (H_{Tmax} - H(s, V))$ is the largest. However, the lowest sum reconstructable value form does not take into account the effects of the photographic distance of the viewpoint, the observation angle and the intersection angle with the existing optimized viewpoints. Therefore, it is guaranteed to have more intersections with the existing optimized viewpoint set by $|S_{v_i} \cap S_V|$; to have the maximum average distance from the existing set of optimized viewpoints by $D_{avr}(v_i, V)$, i.e. to have a larger intersection angle with the existing viewpoints; to have the viewpoint pose facing fronto with the sample point by $\cos(\theta_{avr})$. In addition, the average distance between the chosen viewpoint and its visible sample point set is as close as possible to the preset photographic distance by $|D_{avr}(v_i, S_{v_i}) - D_{gsd}|$. The viewpoint is brought closer to the visible sample point by $D_{avr}(v_i, S_{v_i})$.

The optimal viewpoints are selected by iteration according to Eq. (2), and the final set of optimized viewpoints is obtained when the number of sample points satisfying reconstructability exceeds a certain proportion. The proportion of iteration termination is usually set at 85%-95%, mainly considering that there may be some obvious occluded areas in the scenes and corresponding proxies such as slits that are always unable to satisfy the reconstructability. Although Eq. (2) is formally complex, many of its terms are fixed values. And the time complexity of one iteration can be converted from $O(m\hat{n})$ to $O(\hat{m}\hat{n})$ with a simple algorithmic process, making iteration extremely fast. \hat{m} is the average number of viewpoints visible to a sample point, \hat{n} is the opposite, and m is the number of safe viewpoint points, that is, the cardinality of U . In this way, the viewpoint optimization process is very fast and the number of viewpoints can be determined as required, ensuring coverage of detail without adding too many viewpoints.

Through greedy optimization, the final viewpoint set can be obtained quickly. Then we connect these viewpoints to a flight path based on ACO (Ant Colony Optimization) algorithm with path occlusion considered. Note that we set the cost between two viewpoints based on ideal energy consumption calculated with the constructed mathematical model from UAV flight dynamics.

3. EXPERIMENTS AND DISCUSSION

3.1 Evaluation Metrics

We evaluate our photographic planning method from three aspects, including the scene reconstructability based on reconstruction heuristics, the path quality, and the reconstruction quality of the model. The scene reconstructability represents the reconstructability and corresponding distribution of the sample points. However, it is not the actual quality of the reconstructed model. The path is evaluated with path length, ideal flight time and energy. For model quality, the accuracy and completeness are measured based on ground truth with sampling point clouds on reconstructed models and GT models (Smith et al., 2018; Zhou et al., 2020).

To further verify the effectiveness and advancement of our method proposed, self-evaluation experiments are conducted firstly to verify the effectiveness of our representative sampling and greedy optimization method based on scene reconstructability. Then, in both virtual and real scenes, our method is compared with SOTAs and the method from commercial software.

3.2 Self Evaluation

Effectiveness of Representative Sampling To validate that the sample points generated by our method are more representative to measure the scene reconstructability, multiple sets of controlled experiments are set up. For two types of proxies,

sample points are generated by random, Poisson and our sampling method. The number of sample points of different methods is close to each other. The three optimized viewpoint sets corresponding to three types of sample points are generated using the greedy optimization method described in section 2.2. For the viewpoint set optimized from the sample points of our method, the proportion of sample points of random and Poisson that meet the reconstructability is calculated. Similarly, for the set of optimized viewpoints generated from the other two types of sample points, the proportion of our sample points that meet the reconstructability is calculated. The results are shown in Table 1. The viewpoints (Poisson) represent the viewpoints obtained by optimizing the sample points generated from Poisson sampling, and the same with the other two sampling methods. The optimization iteration terminates when 95% or 90% of the sample points satisfy the reconstructability.

From Table 1, for the viewpoint sets generated from our sample points, the proportion of the other two types of sample points satisfying the reconstructability exceeds the proportion of our sample points, while the proportion of viewpoint sets generated from other types satisfying the reconstructability of our sampling points are all lower than their own proportion. Fig. 6 shows the reconstructability distribution of the sample points obtained by different optimized viewpoints. From Fig. 6 (c) and (f), the reconstructability of the planar regions is similar for different viewpoint sets optimized from different sets of sample points, while it varies more in the edge region. As can be seen from Table 1 and Fig. 6, our sampling method is more representative to measure the reconstructability of the scene, especially in areas such as the edges of the scene.

| Proxy and Scene | | Random Samples | Poisson Samples | Ours Samples |
|-------------------------------|----------------------|----------------|-----------------|--------------|
| School 2.5D Proxy | Viewpoints (Random) | 95.0% | | 93.3% |
| | Viewpoints (Poisson) | | 95.0% | 94.1% |
| | Viewpoints (Ours) | 96.3% | 95.9% | 95.0% |
| Humanities Museum 3D Proxy | Viewpoints (Random) | 90.0% | | 88.2% |
| | Viewpoints (Poisson) | | 90.0% | 89.1% |
| | Viewpoints (Ours) | 90.3% | 89.9% | 90.0% |

Table 1. The proportion of sample points that satisfy the reconstructability with different sets of optimized viewpoints.

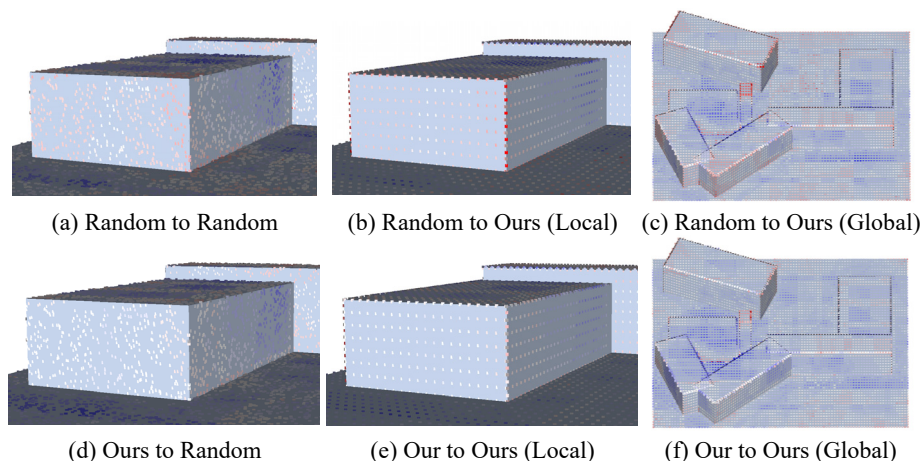


Figure 6. The reconstructability distribution of the sample points obtained by different optimized viewpoints. A to B corresponds to the reconstructability distribution of sample points of B from viewpoints generated by A. The bluer the color of the sample point, the more redundancy exists; the redder it is, the less the reconstructability is likely to be satisfied. When the reconstructable value of a sample point is close to $H_{Tmax}=5.0$, it shows a white color.

Effectiveness of Greedy Optimization To demonstrate the effectiveness of greedy optimization, we generate optimized viewpoint sets that satisfy the reconstructability with 95% and 98% of the sample points on the 2.5D and 3D proxy of the School scene. Then we import the viewpoint sets generated by (Zhou et al., 2020) for comparison, and calculate the proportion of the reconstructed sample points and their distribution, as shown in Fig. 7 and Fig. 8. When using a 2.5D proxy, the sample points and the scene can be basically reconstructed with 270 images based on our method, and the distribution is also uniform. And when 320 images are obtained, the difference with the (Zhou et al., 2020) is small. Similar results are obtained with a 3D proxy. Note that the 330 viewpoints of (Zhou et al., 2020) only satisfy the reconstructability of 96.7% and 96.3% of the sample points, which does not reach the 98% set by our method, proving the effectiveness and efficiency of our greedy method.

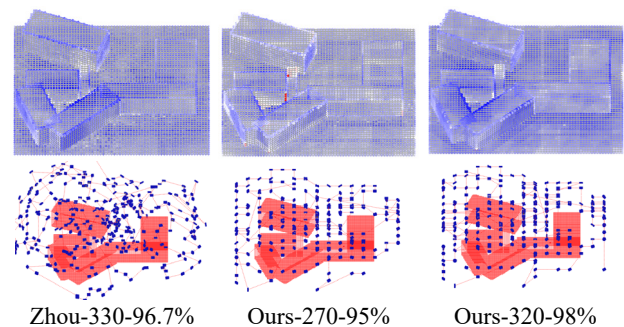


Figure 7. Comparison of different optimized viewpoint sets and the corresponding distribution of sample points (2.5D proxy)

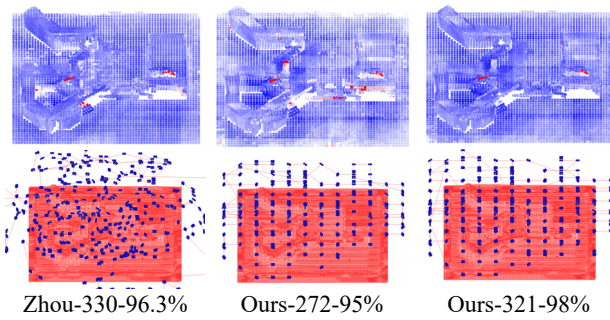


Figure 8. Comparison of different optimized viewpoint sets and the corresponding distribution of sample points (3D proxy)

3.3 Comparison with SOTAs in Virtual Scenes

The UrbanScene3D (Lin et al., 2022) virtual dataset provides proxies and generated photographic paths for methods of (Zhang et al., 2021; Zhou et al., 2020), so we compare our method with them. Two virtual scenes are adopted. The school scene is a sophisticated modern building, while the town scene leans towards a classical style which is considered to be close to cultural heritage. We import the proxies and generate corresponding paths. For image acquisition, we develop a script

based on Unreal Engine 4, which can realize the set of camera parameters, and read the path file to capture high-resolution images automatically. For the model reconstruction, we use Context Capture software to carry out aerial triangulation and dense reconstruction, where the local coordinates of image collection position are all imported to participate aerial triangulation process.

In the processes of photographic planning and image acquisition, the camera parameters consistent with those of (Zhang et al., 2021; Zhou et al., 2020) are set. Table 2 shows various comparisons between the two methods and ours on the School and Town scenes, where the path quality of (Zhou et al., 2020) is not compared due to the unusually longer lengths and times calculated. From Table 2, our method has the best or near-best reconstruction quality on both types of proxies for two scenes. Note that since our path captures multi-angle images in the same position, we get a much shorter length. However, the UAV must hover to change pose, so the energy is not much less but still relatively low since (Zhang et al., 2021) optimize the viewpoint and path simultaneously to get a path with very low energy consumption. Fig.9 shows the visual fidelity of the reconstructed model on the Town scene compared with (Zhou et al., 2020). In all, our method realizes higher quality reconstruction with fewer images and high efficiency of image acquisition.

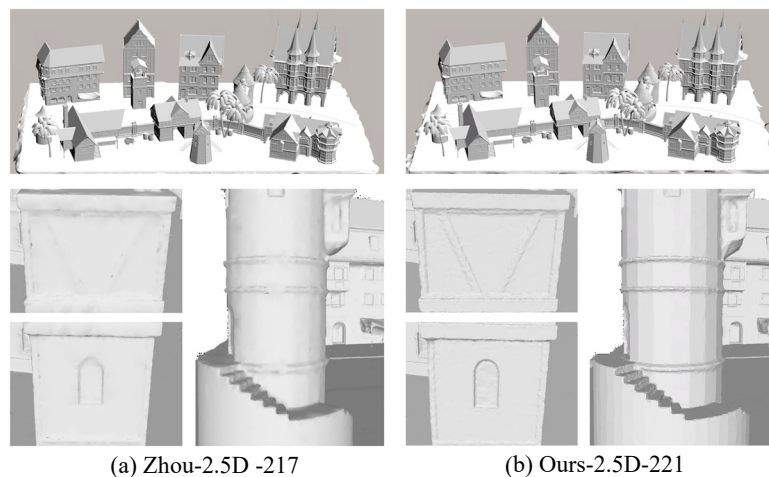


Figure 9. The visual fidelity of the reconstructed model compared with (Zhou et al., 2020)

| Proxy | Method | #Imgs | Length ↓ m | Time ↓ s | Energy ↓ J | RMSE ↓ mm | GSD ↓ mm | Accuracy | | Completeness | | F-Score ↑ % |
|----------------------------|-----------|-------|---------------|-------------|---------------|--------------|-------------|--------------|--------------|--------------|--------------|----------------|
| | | | | | | | | 0.1m ↑ % | 90% ↓ m | 0.1m ↑ % | 90% ↓ m | |
| 2.5D Coarse (School) | Zhang | 330 | 4294 | 1587 | 14013 | 5.97 | 11.98 | 82.96 | 0.325 | 46.64 | 1.893 | 59.71 |
| | Zhou-high | 570 | | | | 4.18 | 9.16 | 83.97 | 0.304 | 51.20 | 1.602 | 63.61 |
| | Zhou-low | 330 | | | | 4.24 | 8.68 | 86.85 | 0.179 | 48.46 | 1.963 | 62.21 |
| | Ours | 320 | 1367 | 1633 | 14532 | 3.92 | 8.26 | 86.93 | 0.182 | 50.10 | 1.704 | 63.57 |
| 3D Inter (School) | Zhang | 330 | 4239 | 1560 | 13776 | 8.70 | 15.96 | 79.95 | 0.413 | 46.81 | 1.853 | 59.05 |
| | Zhou-high | 595 | | | | 3.79 | 8.03 | 15.39 | 1.819 | 9.86 | 2.136 | 12.02 |
| | Zhou-low | 342 | | | | 3.99 | 8.26 | 87.14 | 0.157 | 48.20 | 1.992 | 62.07 |
| | Ours | 321 | 1254 | 1540 | 13787 | 3.65 | 8.25 | 87.63 | 0.162 | 50.72 | 1.723 | 64.25 |
| 2.5D Coarse (Town) | Zhang | 217 | 3457 | 1197 | 10608 | 6.08 | 12.76 | 90.09 | 0.098 | 68.50 | 0.841 | 77.83 |
| | Zhou-high | 511 | | | | 4.02 | 8.77 | 92.01 | 0.061 | 72.02 | 0.721 | 80.80 |
| | Zhou-low | 217 | | | | 4.36 | 8.61 | 92.74 | 0.059 | 70.77 | 0.763 | 80.28 |
| | Ours | 221 | 1098 | 1189 | 10562 | 3.81 | 8.65 | 93.21 | 0.052 | 71.56 | 0.723 | 80.96 |
| 3D Inter (Town) | Zhang | 258 | 3115 | 1212 | 10683 | 6.04 | 12.87 | 89.75 | 0.106 | 68.85 | 0.800 | 77.92 |
| | Zhou-high | 428 | | | | 2.78 | 4.48 | 95.29 | 0.050 | 72.02 | 0.727 | 82.04 |
| | Zhou-low | 259 | | | | 2.46 | 4.11 | 84.32 | 2.070 | 66.75 | 0.926 | 74.51 |
| | Ours | 379 | 1623 | 1544 | 14062 | 2.32 | 4.67 | 95.26 | 0.046 | 72.81 | 0.711 | 82.54 |

Table 2. Qualitative comparison with the state-of-the-art on virtual scenes

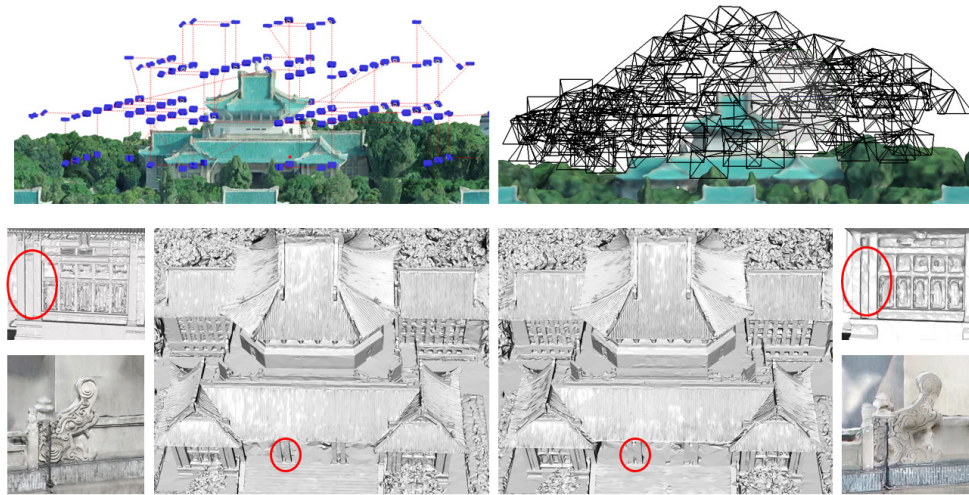


Figure 10. Comparison of the path and reconstructed model between our method (left) and MetaShape software (right)

3.4 Comparison with MetaShape in Real Scene

MetaShape is a reconstruction software that combines 3D reconstruction, path planning and some other functions. It can automatically plan a photographic path based on a delineated 2D plane range and can customize obstacles such as power lines to avoid collisions. And the path generated can be directly exported to DJI P4RTK for flight.

In the real scene, we choose a historic building at Wuhan University to evaluate our method. The building is rather complex, with many intricate details such as ornate carvings on its surface. We compare our method with MetaShape on this scene. The priori proxy was reconstructed from images captured with the straight-down pose at a fixed height. Both methods are set to a ground resolution of 0.6cm, which corresponds to a photography distance of 20 m with DJI P4RTK. Fig.10 shows the comparison of the planned path and the reconstructed model, with the number of viewpoints generated by both methods being close to the same. It can be found that each viewpoint in the path generated by both methods is oriented towards the heritage scene to be reconstructed as far as possible. And both methods obtain a better reconstruction quality of the top of the scene. However, our method gets better visual fidelity of the model in the local detail area.

3.5 Discussion

Our method exploits the geometry structure of the scene and its proxy, which is its most important feature compared to SOATs and MetaShape. First, we generate a set of regularly arranged viewpoints for subsequent optimization. After optimization, these viewpoints remain relatively regularly arranged. Second, we generate SDSM for efficient obstacle avoidance and convenient RTK signal analysis. Third, we realize a more representative measurement of scene reconstructability with sample points generated after geometric planar segmentation. In addition, we adopt a greedy optimization strategy to obtain final optimized viewpoints. However, some improvements can be made to enhance our method, such as using more representative reconstruction heuristics from a learning-based method.

4. CONCLUSIONS

In this paper, we propose an efficient path planning method for digital documentation of cultural heritages, based on which flight

safety and continuity are guaranteed as well as the reconstructability of the whole complex scene. Compared with existing methods, (1) the flight safety and continuity are analyzed and guaranteed based on SDSM; (2) a more representative evaluation of the reconstructability of heritage scene is realized with sampling points on extracted planar regions; (3) the viewpoints are optimized with a fast selection from safe viewpoint set; (4) a regular path is connected to capture images with a shorter length. Compared with the commercial method and the state-of-the-art in both real and virtual scenes, we realize high-quality reconstruction of cultural heritage with higher efficiency in image acquisition, providing a complete and efficient UAV photographic path planning solution for high-quality reconstruction of cultural heritage.

REFERENCES

- Aicardi, I., Chiabrando, F., Maria Lingua, A., Noardo, F., 2018. Recent trends in cultural heritage 3D survey: The photogrammetric computer vision approach. *Journal of Cultural Heritage* 32, 257–266. <https://doi.org/10.1016/j.culher.2017.11.006>
- Bouzas, V., Ledoux, H., Nan, L., 2020. Structure-aware Building Mesh Polygonization. *ISPRS Journal of Photogrammetry and Remote Sensing* 167, 432–442. <https://doi.org/10.1016/j.isprsjprs.2020.07.010>
- Dahlke, D., Linkiewicz, M., Meissner, H., 2015. True 3D building reconstruction: façade, roof and overhang modelling from oblique and vertical aerial imagery. *International Journal of Image and Data Fusion* 6, 314–329. <https://doi.org/10.1080/19479832.2015.1071287>
- El-Hakim, S.F., Beraldin, J.-A., Picard, M., Godin, G., 2004. Detailed 3D reconstruction of large-scale heritage sites with integrated techniques. *IEEE Computer Graphics and Applications* 24, 21–29. <https://doi.org/10.1109/MCG.2004.1318815>
- Furukawa, Y., Hernández, C., 2015. Multi-View Stereo: A Tutorial. *CGV* 9, 1–148. <https://doi.org/10.1561/06000000052>
- Gomes, L., Regina Pereira Bellon, O., Silva, L., 2014. 3D reconstruction methods for digital preservation of cultural heritage: A survey. *Pattern Recognition Letters, Depth Image Analysis* 50, 3–14. <https://doi.org/10.1016/j.patrec.2014.03.023>

- Hepp, B., Nießner, M., Hilliges, O., 2019. Plan3D: Viewpoint and Trajectory Optimization for Aerial Multi-View Stereo Reconstruction. *ACM Trans. Graph.* 38, 1–17. <https://doi.org/10.1145/3233794>
- Koch, T., Körner, M., Fraundorfer, F., 2019. Automatic and Semantically-Aware 3D UAV Flight Planning for Image-Based 3D Reconstruction. *Remote Sensing* 11, 1550. <https://doi.org/10.3390/rs11131550>
- Li, Q., Huang, H., Yu, W., Jiang, S., 2023. Optimized Views Photogrammetry: Precision Analysis and a Large-Scale Case Study in Qingdao. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 16, 1144–1159. <https://doi.org/10.1109/JSTARS.2022.3233359>
- Li, T., Hailes, S., Julier, S., Liu, M., 2017. UAV-based SLAM and 3D reconstruction system, in: 2017 IEEE International Conference on Robotics and Biomimetics (ROBIO). Presented at the 2017 IEEE International Conference on Robotics and Biomimetics (ROBIO), pp. 2496–2501. <https://doi.org/10.1109/ROBIO.2017.8324795>
- Lin, L., Liu, Y., Hu, Y., Yan, X., Xie, K., Huang, H., 2022. Capturing, Reconstructing, and Simulating: The UrbanScene3D Dataset, in: Avidan, S., Brostow, G., Cissé, M., Farinella, G.M., Hasner, T. (Eds.), *Computer Vision – ECCV 2022, Lecture Notes in Computer Science*. Springer Nature Switzerland, Cham, pp. 93–109. https://doi.org/10.1007/978-3-031-20074-8_6
- Liu, X., Ji, Z., Zhou, H., Zhang, Z., Tao, P., Xi, K., Chen, L., Junior, J., 2022. AN OBJECT-ORIENTED UAV 3D PATH PLANNING METHOD APPLIED IN CULTURAL HERITAGE DOCUMENTATION. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences V-1-2022*, 33–40. <https://doi.org/10.5194/isprs-annals-V-1-2022-33-2022>
- Liu, Y., Cui, R., Xie, K., Gong, M., Huang, H., 2021. Aerial path planning for online real-time exploration and offline high-quality reconstruction of large-scale urban scenes. *ACM Trans. Graph.* 40, 1–16. <https://doi.org/10.1145/3478513.3480491>
- Murtiyoso, A., Grussenmeyer, P., 2017. Documentation of heritage buildings using close-range UAV images: dense matching issues, comparison and case studies. *The Photogrammetric Record* 32, 206–229. <https://doi.org/10.1111/phor.12197>
- Nesbit, P.R., Hugenholtz, C.H., 2019. Enhancing UAV–SfM 3D Model Accuracy in High-Relief Landscapes by Incorporating Oblique Images. *Remote Sensing* 11, 239. <https://doi.org/10.3390/rs11030239>
- Nex, F., Remondino, F., 2014. UAV for 3D mapping application: a review. *Appl Geomat* 6, 1–15. <https://doi.org/10.1007/s12518-013-0120-x>
- Pepe, M., Fregonese, L., Scaioni, M., 2018. Planning airborne photogrammetry and remote-sensing missions with modern platforms and sensors. *European Journal of Remote Sensing* 51, 412–436. <https://doi.org/10.1080/22797254.2018.1444945>
- Roberts, M., Shah, S., Dey, D., Truong, A., Sinha, S., Kapoor, A., Hanrahan, P., Joshi, N., 2017. Submodular Trajectory Optimization for Aerial 3D Scanning, in: 2017 IEEE International Conference on Computer Vision (ICCV). Presented at the 2017 IEEE International Conference on Computer Vision (ICCV), IEEE, Venice, pp. 5334–5343. <https://doi.org/10.1109/ICCV.2017.569>
- Schönberger, J.L., Zheng, E., Frahm, J.-M., Pollefeys, M., 2016. Pixelwise View Selection for Unstructured Multi-View Stereo, in: Leibe, B., Matas, J., Sebe, N., Welling, M. (Eds.), *Computer Vision – ECCV 2016, Lecture Notes in Computer Science*. Springer International Publishing, Cham, pp. 501–518. https://doi.org/10.1007/978-3-319-46487-9_31
- Seitz, S.M., Curless, B., Diebel, J., Scharstein, D., Szeliski, R., 2006. A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms, in: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). Presented at the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), pp. 519–528. <https://doi.org/10.1109/CVPR.2006.19>
- Smith, N., Moehrle, N., Goesele, M., Heidrich, W., 2018. Aerial path planning for urban scene reconstruction: a continuous optimization method and benchmark. *ACM Trans. Graph.* 37, 1–15. <https://doi.org/10.1145/3272127.3275010>
- Toschi, I., Ramos, M.M., Nocerino, E., Menna, F., Remondino, F., Moe, K., Poli, D., Legat, K., Fassi, F., 2017. OBLIQUE PHOTOGRAMMETRY SUPPORTING 3D URBAN RECONSTRUCTION OF COMPLEX SCENARIOS. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* XLII-1/W1, 519–526. <https://doi.org/10.5194/isprs-archives-XLII-1-W1-519-2017>
- Yan, F., Xia, E., Li, Z., Zhou, Z., 2021. Sampling-Based Path Planning for High-Quality Aerial 3D Reconstruction of Urban Scenes. *Remote Sensing* 13, 989. <https://doi.org/10.3390/rs13050989>
- Zhang, Han, Yao, Y., Xie, K., Fu, C.-W., Zhang, Hao, Huang, H., 2021. Continuous aerial path planning for 3D urban scene reconstruction. *ACM Trans. Graph.* 40, 1–15. <https://doi.org/10.1145/3478513.3480483>
- Zhou, X., Xie, K., Huang, K., Liu, Y., Zhou, Y., Gong, M., Huang, H., 2020. Offsite aerial path planning for efficient urban scene reconstruction. *ACM Trans. Graph.* 39, 1–16. <https://doi.org/10.1145/3414685.3417791>