

Viewpoint strategies and quality evaluation for 3D reconstruction of exhibition spaces

Yueyang Li, Jiaxuan Zhou*

Guangdong University of Business and Technology, Zhaoqing, China

2317820402@qq.com

Abstract. Digital exhibitions require clear representations of spatial environments and exhibits, and optimized viewpoint selection can help improve reconstruction quality. Using the Library scene as an example, this study compares three viewpoint selection strategies—uniform sequence sampling, farthest-point sampling from the first viewpoint, and farthest-point sampling from a centroid-based starting viewpoint—by reconstructing the scene under identical configurations and evaluating 115 views. The mean Peak Signal-to-Noise Ratio (PSNR) values for the three groups are 22.7395, 22.7906, and 22.9666 dB, respectively, while the corresponding Structural Similarity Index (SSIM) values are 0.72130, 0.72703, and 0.72707. The centroid-based starting group achieves the highest pixel-level consistency, while its structural similarity is close to that of the first-viewpoint group. Accordingly, the centroid-based starting strategy is preferred when prioritizing average image quality.

Keywords: 3D reconstruction, quality evaluation, scene sampling, viewpoint strategy

1. Introduction

The *Opinions on Promoting the Implementation of the National Cultural Digitalization Strategy*, issued in 2022, proposed increasing the supply of digital content for public cultural services and developing digital cultural experiences. The *14th Five-Year Plan for the Protection of Cultural Relics and Scientific and Technological Innovation* incorporated the development of virtual exhibition halls, high-definition acquisition of cultural relics, and their display and utilization into its construction tasks, promoting public access to cultural resources in digital formats. Integrating architectural spaces, artworks, and background information into online environments [1] and expanding access to exhibitions through virtual tours [2] enable visitors to explore exhibition halls and collections beyond on-site visits. The clarity of exhibition layouts and the visibility of exhibit contours and surface details directly determine what visitors can discern from the displayed images. When digital exhibitions are created from image-based reconstruction, investigating the relationship between viewpoint selection and image quality can help reduce redundant observations and allocate a limited number of training images to exhibition areas where details need to be preserved.

1.1. Reconstruction research on virtual exhibitions

Virtual exhibitions integrate architectural spaces, artworks, and historical information through digital documentation. Bent et al. integrated laser-scanned point clouds of the Orsanmichele Church, photogrammetric models of artworks, and historical materials into an online platform, allowing artworks to be viewed within their original architectural context [1]. This form of presentation preserves the spatial relationship between artworks and the building, enabling visitors to choose their own browsing routes, locate exhibits, and examine them. A survey by Li et al. on the online exhibition of architectural art at the Palace Museum found that virtual tours performed well in terms of visual authenticity, although navigation and interaction still required improvement [2]. Digital exhibition halls must therefore not only present spaces and exhibits but also make their contents easy to locate and examine.

When visitors move from browsing an exhibition hall to inspecting individual exhibits at close range, the requirements for rendering contours, textures, and fine details become more demanding, making model size a practical concern during deployment. Because of the excessive size and polygon count of sculpture models, Melendreras Ruiz et al. rescanned and simplified the models to reduce the amount of data that display devices needed to process [3]. Tian et al. supplemented the textures of close-up views of Sanxingdui bronzes in their 3D models with binocular photographs, combining free-viewpoint observation with detailed inspection to capture features that were difficult to represent adequately in the 3D models alone [4]. Khedr et al. employed Gaussian representations to display damaged statues on the web, expanding the range of approaches available for the digital presentation of cultural relics [5]. Exhibition production must address both model performance and the preservation of detail: once visitors can enter a digital scene, they should also be able to discern the shapes and surface characteristics of the exhibits.

Different reconstruction methods utilize image information in different ways. Photogrammetry recovers the geometric surfaces of a scene from correspondences across multiple views and adds material textures. Neural Radiance Fields (NeRF) model spatial density and view-dependent color to synthesize novel-view images [6]. 3D Gaussian Splatting (3DGS) uses explicit Gaussian primitives and visibility-aware rendering to balance appearance optimization with real-time display [7]. When the observation position changes, the visible sides and edges of exhibits, as well as their occlusion relationships, change accordingly. The content and clarity of the input images therefore affect the quality of novel-view rendering. Consequently, in addition to selecting a reconstruction method, it is necessary to determine the positions from which the training images should capture the scene.

Existing studies have improved reconstruction by supplementing scene information or adjusting the optimization process. He et al. used high-precision point clouds to construct synthetic views and reconstructed Gaussian models of cultural relics using geometric and appearance priors [8]. PriNeRF combines adjacent reference views with joint camera optimization to improve novel-view synthesis in urban scenes with sparse views [9]. For images that have already been captured, the distribution of observations can also be adjusted by recombining training viewpoints. When the number of input images remains constant, the capture positions determine which spatial regions can be observed in the images.

Viewpoint selection methods generally assess the value of observations based on observation geometry, image features, and reconstruction feedback. PKS selects keyframes by jointly considering observation geometry and the distribution of valid features [10]. OptiViewNeRF uses uncertainty in an initial reconstruction to select subsequent viewpoints and establishes a maximum–minimum distance baseline for comparison [11]. OUGS initially selects views using the farthest-point sampling rule and then incorporates object masks and uncertainty to add observations, directing subsequent viewpoint selection toward target objects [12]. Assessments of sequence intervals and distances from the scene center can be directly applied to

filtering images with existing camera poses. These two rules produce different combinations of viewpoints, which may lead to differences in the reconstruction of exhibition walls, display platforms, and exhibit details. In their study of the subjective and objective quality of NeRF-synthesized videos, Martin et al. emphasized the importance of pose alignment between reference and synthesized results [13]. If observation conditions differ, the positions and visible extents of objects in the images may also change, making it difficult to attribute differences in evaluation metrics directly to reconstruction quality. Therefore, by fixing the number of input images and the reconstruction configuration and comparing each image against the same set of test views, it is possible to further assess the overall image quality and local detail preservation associated with different combinations of training viewpoints.

1.2. Research questions

Existing studies have improved image quality by introducing geometric priors, optimizing observation positions, and incorporating reconstruction feedback. Uniform sequence sampling selects viewpoints at regular intervals according to the image acquisition order, whereas farthest-point sampling prioritizes positions that are more widely separated in space. The starting point also affects subsequent selections. For digital exhibitions, the effectiveness of viewpoint selection must be evaluated not only by whether overall image quality improves but also by whether details such as display stand contours and exhibit frame edges can be rendered clearly. However, when the number of training images remains constant, the selection of capture positions still requires further investigation. Accordingly, this study establishes three strategies within the same candidate set: uniform sequence sampling (A2), farthest-point sampling starting from the first viewpoint (B2), and farthest-point sampling starting from the centroid-based viewpoint (C2). Their average image quality and visual performance at individual test positions are compared. Specifically, comparisons of A2 with B2 and C2 examine the effects of different sampling rules, while the comparison between B2 and C2 examines the effect of the starting point in farthest-point sampling. These comparisons further explore the relationship between training viewpoint positions and reconstruction quality.

2. Research methods and experimental design

2.1. Experimental methods and design

Both uniform interval sampling and spatially dispersed selection have established methodological foundations. Lin and Lin compared uniform temporal interval sampling, sharpness-based selection, and coverage-sharpness-based selection in a 3DGS workflow for plant phenotyping, demonstrating that, beyond the number of input images, scene coverage and image quality conditions must also be considered [14]. OptiViewNeRF used the max-min distance rule as a baseline for viewpoint selection [11], while OUGS also employed farthest-point selection for initialization [12]. This study uses sequence intervals and distances from the scene center to compare the effects of different input center sets within the same reconstruction pipeline.

The three experimental strategies are designated A2, B2, and C2. A2 represents uniform sequence sampling, B2 represents farthest-point sampling starting from the first viewpoint, and C2 represents farthest-point sampling starting from the centroid-based viewpoint. Three experiments are conducted using training point sets comprising 38 candidate centers. The comparisons between A2 and the two farthest-point sampling strategies evaluate the effects of the sampling rules, while the comparison between B2 and C2 examines the effect of the initial position in farthest-point sampling.

Let the candidate set be $P = \{p_0, \dots, p_{37}\}$, and let the target number of selected points be $K = 24$. A2 includes both endpoints of the sequence and selects points at equal index intervals, with half-integer values

rounded upward (Equation (1)):

$$i_j = \lfloor j(N-1)/(K-1) + 0.5 \rfloor, j = 0, \dots, K-1, N = 38 \quad (1)$$

Equation (1) performs uniform sampling by recording index. For B2, the first point in the candidate sequence serves as the starting point. For C2, the arithmetic centroid c of the 38 three-dimensional centers is first calculated, and the candidate point closest to c is selected as the starting point. The starting point for C2 is determined using Equation (2). Both farthest-point sampling strategies then iteratively add the point farthest from the already selected set, based on the Euclidean distance in XYZ coordinates defined in Equation (3):

$$c = (1/N) \sum_i p_i, p_{start} = \operatorname{argmin}_{p \in P} \|p - c\|_2 \quad (2)$$

$$p_{next} = \operatorname{argmax}_{p \in P \setminus S} \min_{s \in S} \|p - s\|_2 \quad (3)$$

When two candidates have exactly equal distances, the one with the earlier recording time is selected. The starting point for B2 is point 0, while that for C2 is point 21. Each strategy then selects points until 24 centers have been obtained. B2 and C2 use the same distance calculation and iterative selection rules.

A2 and B2 share 15 center positions, as do A2 and C2, while B2 and C2 share 20 centers. The projection method, number of directions per center, camera intrinsic parameters, initialization, number of training iterations, and test views are held constant across all experiments. Each strategy generates 120 training images from 24 centers for comparison. The workflow is illustrated in Figure 1.

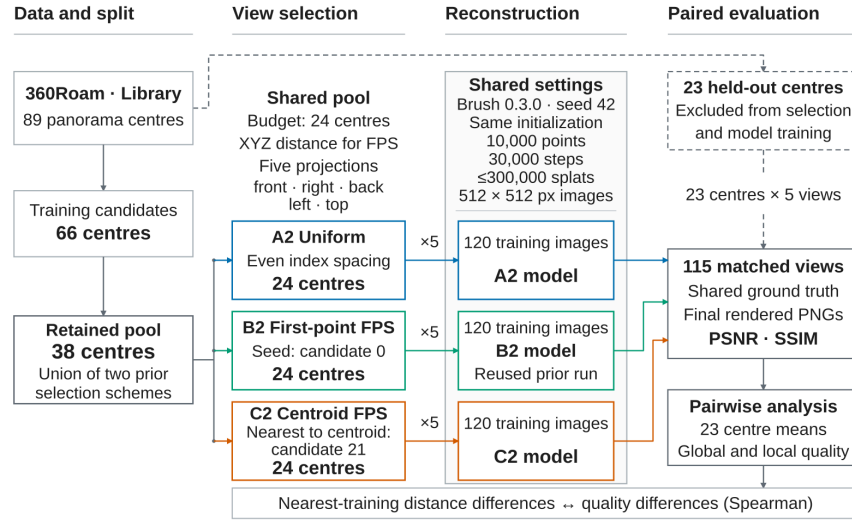


Figure 1. Reconstruction and evaluation workflow for the three viewpoint selection strategies

2.2. Scene data collection

This study uses the Library scene from the publicly available 360Roam dataset. The scene contains exhibition walls, framed artworks, display stands, ceiling fixtures, and carpets, incorporating large indoor surfaces, elongated contours, and repetitive textures. An Insta360 panoramic camera was mounted on a tracked mobile robot to capture images in HDR mode and estimate camera poses. The dataset contains 89 panoramic centers, including 66 candidate training points and 23 test centers. Based on the zero-based indices of the panoramic records, the test positions are 0, 4, 8, ..., and 88. The training records are numbered 0–65, while the test records are numbered 00–22, as shown in Table 1. Figures 2-4 show the training-center distributions for A2, B2, and

C2, respectively. Figures 5 and 6 show the rear-facing reference view at T04 and the forward-facing reference view at T17, respectively.

Table 1. Data collection and training/test configuration

Dataset	Number of Centers	Number of Images
Data acquisition	89	89 panoramas
Data selection	66	66 panoramas
Training data	38	190 perspective images
Uniform sequence sampling (A2)	24	120 perspective images
First-point farthest-point sampling (B2)	24	120 perspective images
Centroid-start farthest-point sampling (C2)	24	120 perspective images
Common test set	23	115 perspective images

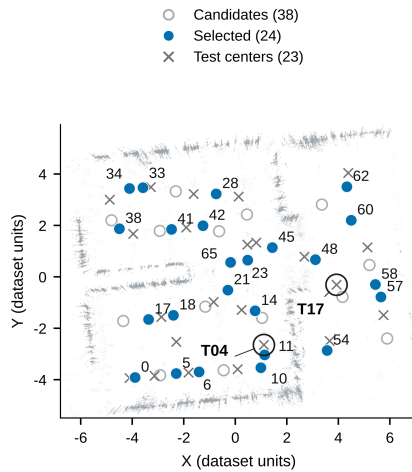


Figure 2. Distribution of training centers for uniform sequence sampling (A2)

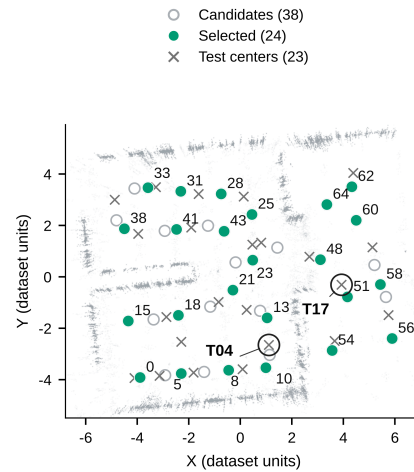


Figure 3. Distribution of training centers for first-point farthest-point sampling (B2)

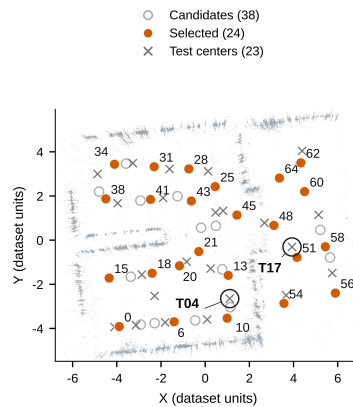


Figure 4. Distribution of training centers for centroid-start farthest-point sampling (C2)



Figure 5. Rear-facing reference view at test position T04

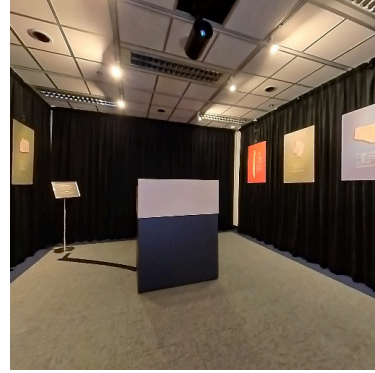


Figure 6. Forward-facing reference view at test position T17

2.3. Image processing and 3D gaussian reconstruction

Using panoramic perspective projection, five directions—front, right, rear, left, and top—were extracted from each center. Each image has a resolution of 512×512 pixels and a field of view of 90° , with focal lengths $f_x = f_y = 256$ and principal point coordinates $c_x = c_y = 256$. The panoramas were resampled using bilinear interpolation and saved as 8-bit RGB PNG images. The rotation for each direction was obtained by composing the original panoramic camera pose with the fixed rotation for the corresponding cube face, while keeping the camera center unchanged. Reconstruction was performed using Brush v0.3.0, running through Vulkan on an NVIDIA GeForce GTX 1660 GPU in a Windows environment. All three groups used the same 10,000 randomly initialized points and their associated colors. The initial coordinates were sampled uniformly within a region centered around the scene centers, with a side length equal to 1.25 times the maximum axis span. The training random seed was fixed at 42, and each experiment ran for 30,000 iterations. The configuration is provided in Table 2.

Table 2. Image processing and training configuration

Dimension	Parameter Settings
Projection resolution and field of view	512×512 px; 90°
Projection directions	Front, right, rear, left, top
Training images per group/common test images	120/115
Initial point count/training seed	10,000/42
Training iterations/maximum number of Gaussians	30,000/300,000
Spherical harmonics degree/densification interval	3/200 iterations
Densification stopping iteration	15,000
SSIM/LPIPS loss weights	0.2/0

2.4. Image quality and spatial difference analysis

Images exported after 30,000 training iterations were paired individually with their corresponding reference images. The full image dimensions, resolution of 512×512 pixels, and three RGB channels were retained, and pixel values were normalized to the interval $[0, 1]$. A higher PSNR indicates a lower pixel-level mean squared

error for the corresponding image. The per-image PSNR values were calculated using Equation (4), and their arithmetic mean was then obtained (Equation (4)):

$$MSE = (1/3HW)\sum_{x,y,c}(I_{xyc} - R_{xyc})^2; PSNR = -10\log_{10}(MSE) \quad (4)$$

Here, H and W denote the image height and width, respectively, while I and R represent the rendered and reference images. A higher SSIM indicates greater consistency with the reference image in terms of local luminance, contrast, and structure [15]. SSIM was calculated using an 11×11 Gaussian window with a standard deviation of 1.5, $C_1 = 0.012$, and $C_2 = 0.032$. Population-based forms were used for local variances and covariance. Zero padding was applied at the boundaries, and the results were averaged across pixels and channels. In addition, a five-pixel border was excluded to calculate valid-window SSIM as a sensitivity check for implementation differences. The five directional views at each test center were first averaged with equal weights. The resulting center-level values were then used to calculate the differences B2-A2, C2-A2, and C2-B2 across the 23 test centers. t-tests were conducted, 95% confidence intervals were calculated, and standardized differences d_z were obtained.

The nearest Euclidean distance from each test center q_i to a training set S was used to characterize the spatial proximity of the training observations: $d_i(S) = \min_{s \in S} \|q_i - s\|_2$. All distances were expressed in the coordinate units of the dataset. The distance advantage of strategy X over strategy Y was defined as $d_i(Y) - d_i(X)$, with positive values indicating that X was closer to the test center. The corresponding quality advantage was defined as $Q_i(X) - Q_i(Y)$, with positive values indicating higher image quality for X . The same direction of comparison was maintained across all three experimental groups, and Spearman correlation was used to characterize the association between changes in spatial proximity and differences in image quality.

3. Results

3.1. Overall quality of the common test views

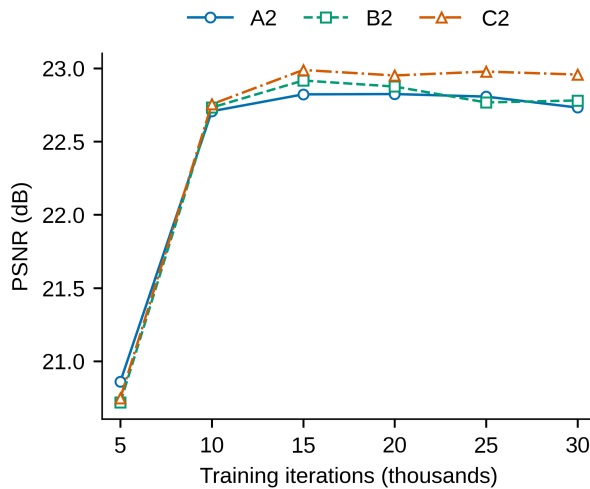


Figure 7. PSNR variation across training checkpoints for the three strategies

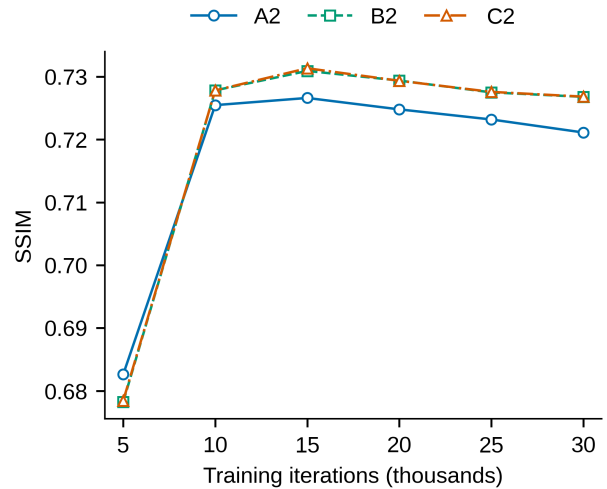


Figure 8. SSIM Variation across training checkpoints for the three strategies

All three groups completed 30,000 training iterations. PSNR increased relatively rapidly between 5,000 and 10,000 iterations in all three groups. SSIM reached its highest recorded checkpoint value at 15,000 iterations and subsequently declined. Among the checkpoints from 10,000 to 30,000 iterations, C2 consistently achieved a higher mean PSNR than A2 and B2. The SSIM curves for B2 and C2 were similar (Figures 7 and 8).

The mean PSNR values for A2, B2, and C2 were 22.7395, 22.7906, and 22.9666 dB, respectively, while their mean SSIM values were 0.721300, 0.727032, and 0.727073, respectively (Table 3). Relative to A2, B2 improved PSNR by 0.0511 dB and increased SSIM by an absolute value of 0.005732. C2 improved the two metrics by 0.2271 dB and 0.005773, respectively. Both farthest-point sampling strategies achieved higher mean pixel-level consistency and structural similarity with the reference images than A2. After changing the starting point for farthest-point sampling, C2 achieved a further PSNR increase of 0.1760 dB over B2, whereas its SSIM increased by only 0.000041. The mean SSIM values of the two groups were therefore very similar.

Table 3. Mean image quality

Strategy	Mean PSNR (dB)	Mean SSIM
A2	22.7395	0.721300
B2	22.7906	0.727032
C2	22.9666	0.727073

Note: PSNR is measured in dB, whereas SSIM is dimensionless.

The center-level paired comparison results are presented in Table 4. The unadjusted p -values for PSNR differences between B2 and A2, C2 and A2, and C2 and B2 were 0.8794, 0.4759, and 0.2853, respectively. The corresponding values for SSIM were 0.4472, 0.4767, and 0.9916. All three pairs of 95% confidence intervals crossed zero. The standardized PSNR difference d_z for C2–B2 was 0.228, whereas the corresponding SSIM effect size was only 0.002, consistent with the similar mean SSIM values of the two groups. After excluding a five-pixel border, the mean SSIM values for A2, B2, and C2 were 0.715340, 0.720980, and 0.721069, respectively, preserving the same ranking.

Table 4. Center-level paired differences in image quality

Comparison	Metric	Mean Difference	95% Confidence Interval	Holm p	d_z
B2–A2	PSNR	0.0511	[-0.6398, 0.7421]	0.9519	0.032
B2–A2	SSIM	0.005732	[-0.009629, 0.021094]	1.0000	0.161
C2–A2	PSNR	0.2271	[-0.4223, 0.8765]	0.9519	0.151
C2–A2	SSIM	0.005773	[-0.010763, 0.022309]	1.0000	0.151
C2–B2	PSNR	0.1760	[-0.1572, 0.5092]	0.8558	0.228
C2–B2	SSIM	0.000041	[-0.007898, 0.007980]	1.0000	0.002

3.2. Differences across test centers

Using A2 as the reference, B2 improved both PSNR and SSIM at eight test centers and reduced both metrics at ten centers. At the remaining five centers, the two metrics changed in opposite directions. For C2, the corresponding counts were 10, 11, and 2 centers, respectively. C2 therefore improved both metrics at two more centers than B2, although the locations of these improvements did not completely coincide. In the direct comparison between B2 and C2, C2 achieved higher values for both metrics at nine centers and lower values

for both at seven centers. At the remaining seven centers, each strategy performed better on one of the two metrics (Figures 9 and 10). These differences indicate that changing the starting point for farthest-point sampling redistributes local image-quality advantages, making it necessary to compare the reconstruction results at specific exhibition locations.

Compared with A2, C2 exhibited PSNR differences ranging from -2.0345 to 4.1192 dB across the test centers. The median difference was only 0.0391 dB, lower than the mean difference of 0.2271 dB. The median SSIM difference was -0.00357 , whereas the mean difference was 0.005773 . These results indicate that C2's advantage in mean image quality was driven primarily by substantial improvements at certain locations. Although B2 and C2 shared 20 training centers, replacing the remaining four centers resulted in different advantages at different test positions. This finding suggests that the starting point for farthest-point sampling affects the final set of selected positions and, consequently, local image quality. The advantage of C2 is therefore primarily reflected in its higher overall mean quality.

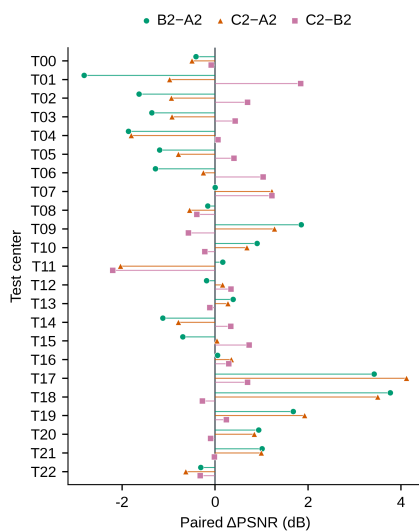


Figure 9. Paired differences in pSNR

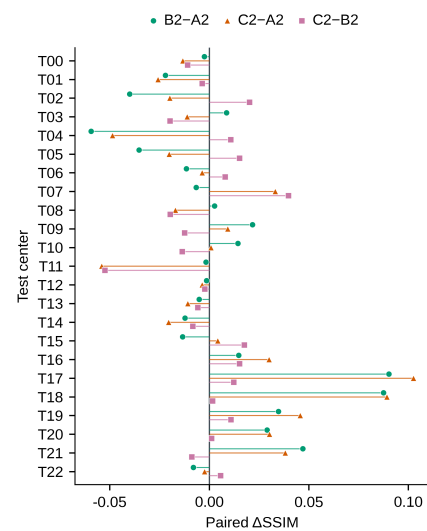


Figure 10. Paired differences in SSIM

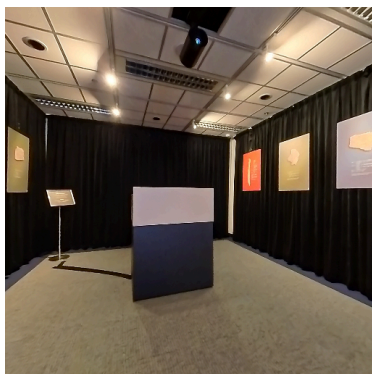


Figure 11. Forward-facing reference image at T17

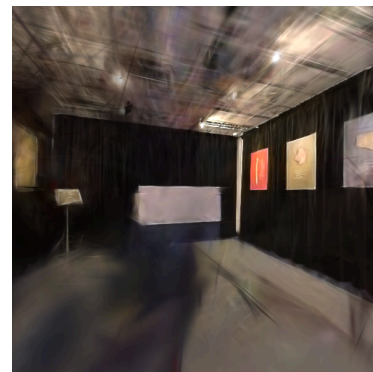


Figure 12. Forward-facing reconstruction by A2 at T17

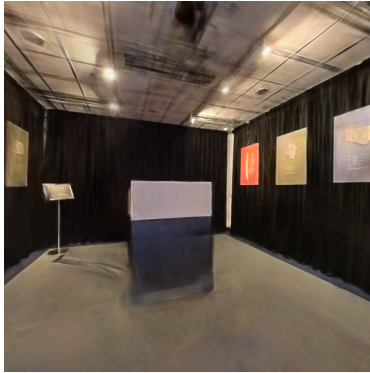


Figure 13. Forward-facing reconstruction by B2 at T17

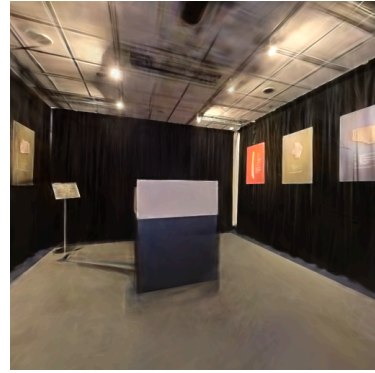


Figure 14. Forward-facing reconstruction by C2 at T17



Figure 15. Rear-facing reference image at T04

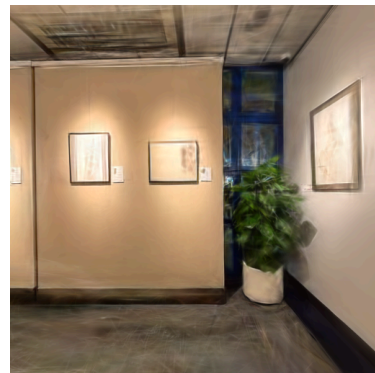


Figure 16. Rear-facing reconstruction by A2 at T04



Figure 17. Rear-facing reconstruction by B2 at T04

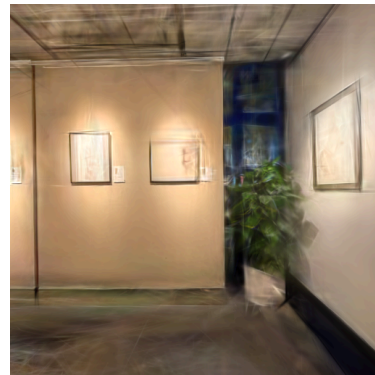


Figure 18. Rear-facing reconstruction by C2 at T04

Figures 11–18 compare the reconstruction results at locations containing display stands and exhibit frames. In the forward-facing view at test center 17, A2 exhibited noticeable ghosting along the display stand contours, in the ceiling grid, and in the carpet textures. B2 and C2 produced clearer display stand edges, with object contours more closely resembling those in the reference image. The PSNR values for the three groups were 15.951, 22.477, and 22.663 dB, respectively, while the corresponding SSIM values were 0.62160, 0.75943,

and 0.76633. The visual improvements were consistent with increases in both metrics. C2 achieved the highest values for both metrics, although the carpet textures remained blurred in both B2 and C2. The improvements were primarily reflected in the improved discernibility of the display stand contours.

In the rear-facing view at test center 04, the exhibit frame edges were clearer in A2, whereas B2 and C2 exhibited duplicated edges. The PSNR values for the three groups were 23.725, 21.900, and 20.957 dB, respectively, and the corresponding SSIM values were 0.73017, 0.66988, and 0.67541. A2 outperformed both other strategies on both metrics. These two comparisons show that farthest-point sampling achieved better contour reconstruction at the display stand location, whereas uniform sequence sampling preserved clearer edges at the exhibit frame location.

3.3. Relationship between training viewpoint proximity and quality differences

Table 5 compares the distances from each test center to its nearest training center. The mean distances for A2, B2, and C2 were 0.6680, 0.6731, and 0.6919, respectively, while the corresponding medians were 0.6030, 0.6030, and 0.7019. A2 had the smallest mean distance, but its maximum distance reached 1.2750. Both B2 and C2 reduced the maximum distance to 1.2272, a decrease of 3.75% relative to A2. These results indicate that uniform sequence sampling brought test positions closer to training positions on average, whereas the two farthest-point sampling strategies reduced the gap between the most distant test positions and the training sets. Although the two farthest-point sampling strategies achieved the same maximum distance, both the mean and median distances for B2 were smaller than those for C2, indicating that changing the starting point still affected the spatial proximity between training and test positions.

Table 5. Minimum distances from test centers to training centers

Statistic	A2	B2	C2
Mean	0.6680	0.6731	0.6919
Median	0.6030	0.6030	0.7019
Maximum	1.2750	1.2272	1.2272

For the three pairwise comparisons—B2–A2, C2–A2, and C2–B2—the Spearman correlation coefficients between distance advantages and PSNR differences were 0.382, 0.583, and 0.534, respectively. The corresponding coefficients for SSIM differences were 0.530, 0.561, and 0.694 (Figures 19 and 20). In other words, at a given test position, the strategy whose training centers were closer generally achieved better image quality. The strongest association between distance advantage and SSIM difference occurred in the comparison between C2 and B2. This finding suggests that the positional differences resulting from the two starting-point configurations were closely associated with changes in local structural similarity, indicating a positive relationship between training viewpoint proximity and reconstruction quality.

Adjustments to viewpoint selection changed the distances from only some test positions to their nearest training centers, and positions with larger distance changes exhibited corresponding differences in image quality. At test center 17, the distance from C2 to its nearest training center decreased from 1.2750 under A2 to 0.5242, while PSNR increased by 4.1192 dB. The display stand contours in the forward-facing view also became clearer. At test center 04, this distance increased from 0.3887 to 0.8943, while PSNR decreased by 1.8029 dB, and duplicated edges appeared around the exhibit frame in the rear-facing view (Figures 11–18).

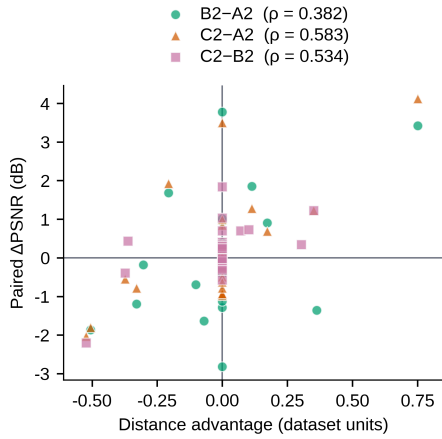


Figure 19. Correlation between differences in nearest training distance and differences in PSNR

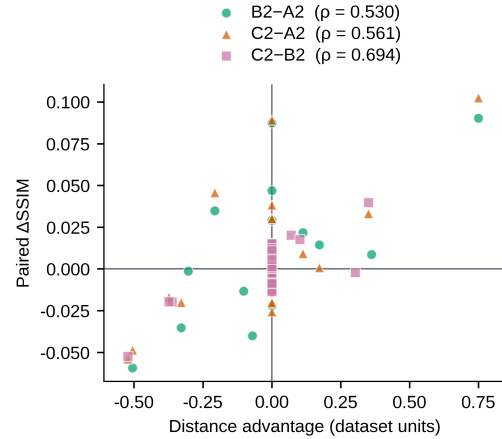


Figure 20. Correlation between differences in nearest training distance and differences in SSIM

4. Discussion

4.1. Viewpoint selection strategies

Under conditions in which the number of training images was held constant, centroid-start farthest-point sampling achieved the highest mean image quality among the three strategies. Compared with uniform sequence sampling (A2), C2 increased the mean PSNR by 0.2271 dB and the mean SSIM by 0.005773. Compared with first-point farthest-point sampling (B2), C2 increased the mean PSNR by 0.1760 dB, while the mean SSIM values remained similar. Accordingly, centroid-start farthest-point sampling (C2) can be prioritized to achieve better average pixel-level consistency and structural similarity with the same number of input images.

4.2. Observation distribution and reconstruction quality

Farthest-point sampling improved the proximity of training observations to relatively sparsely observed regions. The maximum nearest-neighbor distances for both B2 and C2 were 3.75% lower than that for A2, whereas their mean nearest-neighbor distances increased slightly. This reflects the fact that uniform sequence sampling achieved a smaller mean distance, while farthest-point sampling prioritized reducing the largest observation gaps. At test center 17, C2 reduced the distance to the nearest training center, increased PSNR, and produced clearer display stand contours. This finding supports the positive association between distance advantages and image-quality differences: the quality improvements associated with spatially dispersed sampling were specifically manifested in shorter distances to training observations at certain exhibition locations.

OptiViewNeRF selects subsequent viewpoints based on uncertainty in an initial reconstruction [11], while OUGS supplements farthest-point initialization with object masks and uncertainty to add observations [12]. Both methods adjust the allocation of observations using reconstruction feedback. In contrast, under a fixed input budget, C2 achieved both a smaller maximum nearest-neighbor distance and higher mean image quality in this study. This suggests that, for exhibition imagery, centroid-start farthest-point sampling can balance the

proximity of training observations to the most sparsely observed test positions with the overall image quality of the reconstructed results. Therefore, for viewpoint selection from existing images, C2 can be prioritized to improve observation coverage and achieve better overall reconstruction quality.

5. Conclusion

Under identical training image counts and reconstruction configurations, this study compared three viewpoint selection strategies: uniform sequence sampling, first-point farthest-point sampling, and centroid-start farthest-point sampling. C2, the centroid-start farthest-point strategy, achieved the highest mean PSNR, exceeding the other two groups by 0.2271 and 0.1760 dB, respectively. Its mean SSIM was higher than that of the uniform sequence strategy and similar to that of the first-point farthest-point strategy. Its advantage was primarily reflected in improved average pixel-level consistency. Therefore, for the exhibition imagery examined in this study, prioritizing a viewpoint selection strategy that starts from the candidate point closest to the centroid and iteratively adds the farthest remaining candidate can produce better overall image quality with the same number of input images.

The distance advantage of training positions was positively associated with local image-quality improvements, while the different combinations of positions resulting from changes in the starting point corresponded to different distributions of image quality. Viewpoint selection optimization involves not only the spatial distribution of training positions but also the effects of specific position combinations on the reconstruction of exhibit details. In practice, the centroid-start farthest-point strategy can be used to organize training images, with reconstruction results selected according to mean quality across common test views. Local display quality should then be assessed by examining the contours, textures, and ghosting of key exhibits. Future research could further incorporate exhibit visibility and image features to investigate how complementary observations affect local image quality and to optimize image selection for prominent exhibition objects.

Funding

The 2026 Key Research Planning Project of Guangdong University of Business and Technology, entitled "Research on Quality Evaluation and Acquisition Strategy Optimization for Campus Space Reconstruction Based on 3D Gaussian Splatting: A Case Study of Guangdong University of Business and Technology" (KYZ2026006)

References

- [1] Bent, G. R., Pfaff, D., Brooks, M., Radpour, R., & Delaney, J. (2022). A practical workflow for the 3D reconstruction of complex historic sites and their decorative interiors: Florence as it was and the church of Orsanmichele. *Heritage Science*, 10, Article 118. <https://doi.org/10.1186/s40494-022-00750-1>
- [2] Li, J., Nie, J.-W., & Ye, J. (2022). Evaluation of virtual tour in an online museum: Exhibition of architecture of the Forbidden City. *PLOS ONE*, 17(1), e0261607. <https://doi.org/10.1371/journal.pone.0261607>
- [3] Melendreras Ruiz, R., Sánchez Allegue, P., Marín Torres, M. T., Pardo Ríos, M., Cerón Madrigal, J. J., & Escribano Tortosa, D. (2024). Analysis of the user experience (on site vs. virtual reality) through biological markers and cognitive tests in museums: The case of Museo Cristo de la Sangre (Murcia, Spain). *Virtual Reality*, 28, Article 48. <https://doi.org/10.1007/s10055-023-00928-3>

- [4] Tian, F., Li, K., Huang, X., Zhang, X., Wang, N., Song, Y., Zhu, Q., & Li, Y. (2024). An empirical study of virtual museum based on dual-mode mixed visualization: The Sanxingdui bronzes. *Heritage Science*, 12, Article 146. <https://doi.org/10.1186/s40494-024-01241-1>
- [5] Khedr, M., Metawie, M., Eldin, A. S., Hamdy, A., El-Shihy, A., Azab, S., Abdelaty, A., & Marzouk, M. (2025). 3D visualization of damaged statues using Gaussian splatting and web interface integration. *npj Heritage Science*, 13, Article 587. <https://doi.org/10.1038/s40494-025-02063-5>
- [6] Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., & Ng, R. (2022). NeRF: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1), 99–106. <https://doi.org/10.1145/3503250>
- [7] Kerbl, B., Kopanas, G., Leimkühler, T., & Drettakis, G. (2023). 3D Gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4), Article 139, 1–14. <https://doi.org/10.1145/3592433>
- [8] He, Y., Zhang, X., Xie, Z., Li, D., He, J., & Chen, R. (2026). A dual-prior driven Gaussian splatting framework for high-fidelity reconstruction of museum artifacts. *npj Heritage Science*, 14, Article 69. <https://doi.org/10.1038/s40494-026-02330-z>
- [9] Chen, K., Dong, B., Wang, Z., Cheng, P., Yan, M., Sun, X., Weinmann, M., & Weinmann, M. (2024). PriNeRF: Prior constrained Neural Radiance Field for robust novel view synthesis of urban scenes with fewer views. *ISPRS Journal of Photogrammetry and Remote Sensing*, 215, 383–399. <https://doi.org/10.1016/j.isprsjprs.2024.07.015>
- [10] Azimi, A., Hosseininaveh Ahmadabadian, A., & Remondino, F. (2022). PKS: A photogrammetric key-frame selection method for visual-inertial systems built on ORB-SLAM3. *ISPRS Journal of Photogrammetry and Remote Sensing*, 191, 18–32. <https://doi.org/10.1016/j.isprsjprs.2022.07.003>
- [11] Li, Y., Li, R., Li, Z., Guo, R., & Tang, S. (2025). OptiViewNeRF: Optimizing 3D reconstruction via batch view selection and scene uncertainty in neural radiance fields. *International Journal of Applied Earth Observation and Geoinformation*, 136, Article 104306. <https://doi.org/10.1016/j.jag.2024.104306>
- [12] Li, H., Chen, Q., Kalkofen, D., & Chen, H.-T. (2026). OUGS: Active view selection via object-aware uncertainty estimation in 3DGS. *Computer Graphics Forum*, 45(2), e70363. <https://doi.org/10.1111/cgf.70363>
- [13] Martin, P., Rodrigues, A., Ascenso, J., & Queluz, M. P. (2025). NeRF view synthesis: Subjective quality assessment and objective metrics evaluation. *IEEE Access*, 13, 26–41. <https://doi.org/10.1109/ACCESS.2024.3522768>
- [14] Lin, J.-H., & Lin, T.-T. (2026). From 3DGS scenes to plant traits: A scalable extraction and segmentation framework for muskmelon phenotyping. *Frontiers in Plant Science*, 17, Article 1783465. <https://doi.org/10.3389/fpls.2026.1783465>
- [15] Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4), 600–612. <https://doi.org/10.1109/TIP.2003.819861>