Research on Cultural Relic Restoration and Digital Presentation Based on 3D Reconstruction MVS Algorithm: A Case Study of Mogao Grottoes' Cave 285

Mengyao Gao

Information Engineering Department, Southwest Jiaotong University Hope College, Chengdu 610000, China

1807050220@stu.hrbust.edu.cn

Abstract. This paper delves into the realm of three-dimensional (3D) reconstruction technology, specifically examining the principles underlying Multi-View Stereo (MVS) techniques, encompassing pose calculation, dense reconstruction, surface reconstruction, and texture mapping. It scrutinizes the application of these methodologies utilizing Colmap and Open Multi-View Stereo reconstruction library (OpenMVS) within Cave 285 of the Mogao Caves in Dunhuang, culminating in a meticulous technical evaluation and analysis. The primary objective is to deliberate upon the non-invasive, multi-angle accurate reconstruction of the cave, aimed at furnishing archaeologists with the utmost detailed information concerning its features, thereby facilitating the planning and execution of artifact restoration endeavors. This investigation discerns that MVS 3D reconstruction technology affords highly precise, multi-perspective models, while concurrently mitigating any potential secondary harm to the artifacts. Moreover, it elucidates the intricate contours of the artifacts, documents historical damage, and engenders comprehensive three-dimensional datasets for digital preservation. Subsequent research endeavors could continue to innovate and amalgamate enhancements, furnishing three-dimensional databases to diverse cultural heritage preservation entities for the enduring safeguarding of artifacts. Integration of technologies such as Convolutional Neural Networks holds promise for future applications, potentially contributing to the advancement of research and development within the metaverse.

Keywords: Mogao Caves, Multi-View Stereo, OpenMVS, Artifact Restoration

1 Introduction

The Mogao Grottoes in Dunhuang are among the largest, longest-standing, and relatively well-preserved Buddhist historical relics in China, representing the zenith of Chinese Buddhist cultural and artistic achievements from the 4th to the 14th century. They are a crystallization of ancient Chinese multi-ethnic and Eurasian cultures on the Silk Road [1]. Not only are they a culmination of ancient Chinese art, but they also carry a millennium of historical and cultural heritage. The murals, sculptures, and other

Y. Wang (ed.), *Proceedings of the 2024 2nd International Conference on Image, Algorithms and Artificial Intelligence (ICIAAI 2024)*, Advances in Computer Science Research 115,
https://doi.org/10.2991/978-94-6463-540-9_76

artworks within the caves are vital historical materials for studying the politics, economy, culture, and religion of ancient China and even the Eurasian continent. However, due to natural and human factors, the artifacts in the Mogao Grottoes face severe damage, making digital preservation especially important [2]. Digital preservation not only records the current state of the artifacts, ensuring their permanent preservation, but also allows for the repair and reproduction of the caves through digital technology, which is of great significance for the inheritance and promotion of Dunhuang culture. Currently, digital restoration of cultural relics has gradually become popular worldwide. In the past two years, China has also applied multiscale point cloud and surface texture technologies to the restoration of colored ceramics of the Terracotta Warriors in Xi'an, and has used point cloud technology for the research on the conservation of lacquerware. In Europe, 3D scanning technology has been used for the digital restoration of the Parthenon Temple, and in Italy, 3D scanning was utilized to repair several artifacts in L'Aquila.

This study is based on the Multi-View Stereo (MVS) algorithm for dense reconstruction, converting multi-view image data into 3D point clouds and using surface reconstruction to generate visualized 3D surface models. This allows for high-precision reconstruction of the 3D model of Cave 285 in the Mogao Grottoes and discusses the optimization of the model through point cloud fusion, finally improving visual effects with texture mapping. This reconstruction process can accurately capture the spatial structure and detailed features of Cave 285 and demonstrate the efficiency of the MVS algorithm in identifying and assessing damaged areas. The application of this technology allows archaeologists to observe and measure artifacts in detail without direct contact, accurately identifying damaged parts and the extent of damage, providing a solid basis for future artifact restoration work in the Mogao Grottoes.

2 Literature Review

Since the discovery of the Scripture Cave in the Mogao Caves in Dunhuang in 1900, more than 700 caves have been found to date, encompassing over 45,000 square meters of murals and more than 50,000 scrolls and documents. Spanning from the 4th to the 14th century, due to looting in the earlier years, human damage caused by tourists in recent years, and the natural challenges of desert location, including geological activity and sand erosion, the Mogao Caves face significant threats.

To ensure the sustainable development of the Mogao Caves' ecosystem, Fan Jinshi (2004), proposed a multidisciplinary comprehensive archaeological approach. This approach explores the application of computer and digital technologies in the restoration and protection of cave murals, aiming to construct a complete digital storage and reproduction system for the relics [3]. Additionally, the Artificial Intelligence Research Institute of Zhejiang University (2003) developed an intelligent Dunhuang cultural relic tracing aid system [4]. This system utilizes image segmentation and machine learning technologies to simulate mural evolution, achieving virtual color restoration and the virtual evolution and repair of diseases. This research was quite innovative given the technological context of the time. However, on one hand, due to the immaturity of the technology, and on the other hand, scholars' emphasis on the application in two-dimensional scenes, the study still falls short in supporting the restoration of Dunhuang artifacts due to the limitations of the two-dimensional scene perspective.

In the following decade, most research focused on environmental monitoring and geological exploration of the Mogao Caves, with virtually no progress in digital restoration and protection of cultural relics. It was not until Tencent and the Dunhuang Research Institute collaborated to establish the Cultural Heritage Digital Creative Technology Joint Laboratory that archaeological technology and digital relic restoration once again attracted widespread scholarly interest. In 2022, YU Tianxiu and others applied computer vision, convolutional neural networks (CNNs), and generative adversarial networks (GANs) for the classification and comparison of Dunhuang murals, feature detection, color restoration, identification of damaged and deteriorated areas, and analysis of influencing factors to ultimately develop a virtual restoration model [5]. However, this study still focuses more on the restoration and protection of the two-dimensional plane of the Mogao Caves murals.

For the MVS, Hailin Jin and others (2005) utilize the 3D shape and complex appearance of a scene from multiple calibrated views under fixed illumination, thereby generating non-Lambertian views through 3D reconstruction [6]. Philippe Pons et al. (2007) proposed a novel approach for MVS algorithm and multiple video sequences, employing a global measure of image similarity, selected based on imaging conditions and scene characteristics, and addressing the issues of projective distortion and occlusions in the model [7]. Then, Korbinian Schmid and others (2011) explored an autonomous data collection algorithm for MVS based on drones, aiming at planning and reconstructing systems for outdoor sites [8]. It is evident that as scholars have developed the foundational theory of MVS algorithms, most applied research has concentrated on geological terrain exploration and medical imaging. In 2019, Nobuaki Kuchitsu and others conducted a simple 3D measurement and analysis of the Wareishi Rock Cliff Sculpture and their replicas, using multi-view stereophony, thereby assessing the degradation status of the heritage and aiming at its preservation [9].

Consequently, there has been less scholarly focus on the application of MVS algorithms in the conservation of cultural heritage. The concepts of scientific archaeology and relic restoration have leaned more towards the use of technologies such as artificial intelligence and deep learning. Therefore, this study will focus on this research gap, proposing the significant role of MVS algorithms in the digital preservation of the Mogao Caves in Dunhuang, and exploring the implementation of the technology and its future application possibilities.

3 3D Reconstruction of Cave 285 in the Mogao Grottoes

3D reconstruction technology can effectively promote the advancement and perfection of artifact restoration and digital preservation. Since Cave 285 of the Mogao Caves in Dunhuang is the earliest excavated cave with precise historical records, its murals are the most severely corroded and damaged, and it is essentially no longer open to the public. Therefore, this section will elaborate on the collection and preprocessing of the image dataset for Cave 285 based on 3D reconstruction MVS technology, perform feature matching and point cloud acquisition of the main chamber mural atlas, produce a 3D model, and ultimately optimize it with texture mapping. This approach restores the main chamber murals and sculptures from multiple viewpoints, thus perfecting the digital archive of the artifacts.

3.1 Dataset for Reconstruction of Cava 285

High-quality image collection is one of the key tasks in this part. It is important to choose appropriate lighting conditions to avoid overexposure caused by strong direct light or image distortion due to overly dim environments. Most importantly, the mural image dataset should be captured from high-resolution cameras and multiple viewpoints to ensure 60% to 70% overlap of the images, which guarantees accurate feature point matching during experiments, ultimately leading to a more precise 3D model. Based on the analysis above, to ensure the authenticity and accuracy of the dataset, the data for this study were collected from the official "Digital Dunhuang" information platform.

Subsequently, in the image preprocessing work, viewpoint correction is performed on the dataset. Images taken with wide-angle lenses may exhibit distortion, causing the edges to bend, and to ensure the correctness of visual perspective in the images, distortion correction is carried out during preprocessing. Furthermore, images shot under varying lighting conditions can exhibit significant overexposure and color imbalance, necessitating adjustments in hue, saturation, and exposure compensation to ensure visual consistency in color, adjusting to ensure both dark and bright detail areas are properly displayed, thus accurately presenting the image collection. For images with low clarity, it is also necessary to enhance details using sharpening filters and reduce noise to improve quality.

3.2 Methodology

Based on the MVS algorithm, 3D reconstruction first involves inputting source data and calculating pose estimation, which estimates the corresponding three-dimensional scene positions from two-dimensional images, thereby determining the spatial positioning and orientation of each image. This step primarily relies on Colmap technology. Next, a dense reconstruction is performed using the Semi-Global Block Matching (SGM) algorithm. This phase begins with image matching using neighborhood frames after irrelevant data is discarded. The neighborhood frame selection strategy considers the common viewpoint f, the angle w_N between two images V, R under different angles of the common viewpoint f, the similarity of resolution in the two images as w_s, and the minimum covered area in the two images as $area$. The score is calculated as follows:

$$Score\ (V) = area_R \cdot \sum_{f \in F_V \cap F_R} w_N(f) \cdot w_s(f) \tag{1}$$

$$w_N(f) = \prod_{\substack{V_i, V_j \in N \\ s.t.\ i \neq j, f \in F_{V_i} \cap F_{V_j}}} \mathrm{w}_\alpha(f, V_i, V_j) \quad (2)$$

This part requires optimization of the neighborhood frame options. The basic logic involves selecting a label for each node to minimize the overall energy, which involves the use of Markov Random Field energy functions. Subsequently, depth maps are initialized using the Delaunay method, followed by cost aggregation based on the census transform to achieve two-dimensional optimization. The formula is constructed as follows10:

$$E(D) = \sum_{x_b} C\left(x_b, D(x_b)\right) + \sum_{x_N} P_1 T[D(x_b) - D(x_N) = 1] + \sum_{x_N} P_2 T[D(x_b) - D(x_N) > 1] \quad (3)$$

The matching cost, C, with the second and third terms being smoothing terms, thus achieving the desired continuity of disparity. Afterward, a consistency check of the disparity is performed, followed by disparity optimization.

Next, using a tetrahedral mesh division and corresponding weight calculations, the surfaces are extracted, achieving surface reconstruction to obtain the basic mesh model. This is followed by minimizing the objective function defined on the mesh model to optimize the position of each vertex vi. The general framework is as follows11:

$$E(\{v_i\}) = E_p(\{v_i\}) + E_r(\{v_i\}) + [E_s(\{v_i\})] \quad (4)$$

The data term $\mathrm{E_p(\{v_i\})}$ is based on photometric consistency measurements, the regularization term $\mathrm{E_r(\{v_i\})}$ ensures the smoothness of the mesh model, and $\mathrm{E_s(\{v_i\})}$ is for contour consistency measurement. Finally, after implementing the above calculations, a Markov Random Field is used to select the face view, and global texture color correction is performed. It is possible to consider using a Poisson distribution for local blending to create the color information of the object, achieving the final output of the 3D model. The above algorithm is implemented using Open Multi-View Stereo reconstruction library (OpenMVS).

4 Experiment for Cave 285 in the Mogao Grottoes

4.1 Experiment Process

This study utilizes both Colmap and OpenMVS, two open-source frameworks, for constructing Cave 285 of the Mogao Caves. Colmap offers three functionalities: Structure from Motion (SFM) sparse modeling, densify, and mesh. However, due to the poor quality of the mesh model obtained from Colmap, the reconstruction experiment also employs the OpenMVS framework to optimize and present the best visual effects. Since direct modeling with OpenMVS results in an excessively large amount of point clouds, to reduce the spatial and temporal complexity of the experiment and the high memory demands.

Firstly, to ensure the quality of the acquired point cloud, sparse point clouds are generated using SFM technology in Colmap based on the pose calculations of the source image set, followed to obtain depth maps. These are then transferred to the OpenMVS interface, where the SGM algorithm is used for matching cost calculation and cost aggregation, ultimately leading to disparity optimization. Furthermore, the initial mesh model obtained is subjected to surface reconstruction using the Delaunay triangulation method, minimizing sharp angles in the mesh model and preserving the original shape of the point cloud. Lastly, after optimizing the model, texture mapping is performed to display the overall color information of the main chamber.

4.2 The Analysis of MVS Reconstruction Results

Following the experimental steps, after feature extraction and matching, we obtained the results of the sparse reconstruction point cloud model (figure 1). Then, by inputting the sparse reconstruction results into OpenMVS for optimization, we obtained the dense reconstruction point cloud (figure 2). From figure 1, it is clear that the sparse point cloud provides an initial framework for the entire 3D reconstruction process, offering shooting angles and camera paths. The point cloud is mainly concentrated in the center of the model, with fewer and more sparse distributions on both sides, and almost no point cloud appearance in the upper and lower beam parts of the model. The results show that the point cloud distribution across the model is uneven, with the middle position achieving more successful extraction and matching, and the peripheral areas not covered. This result indicates that subsequent optimizations need more accurate and comprehensive perspective data for compensation.

Through figure 2, it is evident that the dense reconstruction results in a more densely packed point cloud, extracting more details for matching, allowing for a clearer visualization of the colors and contours of the cave murals. The sculptures and carvings on the correct side of the cave can be more accurately restored, and the mural colors do not show overexposure, fading, or mottling, appearing more complete and rich visually. However, due to the lack of some perspective information in the sparse reconstruction, the dense point cloud results are mainly focused on the main niche and the left side of the main Buddha statue, with the entire main chamber's dome part and the right side area failing to be presented successfully. The main reason the dome mural details were not successfully reconstructed is that the photo dataset did not capture too many details, and there were missing shooting angles, which need further improvement in future research. Looking at both parts of the results as a whole, from the initial framework to dense reconstruction for texture refinement, the overlap between the two is relatively high, without significant errors and deviations, and the successfully reconstructed areas are relatively close to the actual scene. Therefore, it can be seen that the 3D reconstruction of the cave using MVS technology can well restore and present the complex details of cultural relics, reconstructing the cave while ensuring its feature contours and restoring the real color details.

In summary, the implementation of MVS technology through reconstruction can effectively restore the detailed color features of Cave 285, achieving a model close to the real scene with high accuracy in the successfully presented sections, providing extensive information on the murals and sculptures. However, due to insufficient coverage of angles and lack of significant perspective changes in the image dataset, texture information was not correctly extracted and mapped from sparse to dense reconstruction; also, due to the limitations of the experimental environment, the current software capabilities and computing power available for this study are limited, resulting in poor final texture reconstruction. In subsequent research, the shooting angles and positions of the dataset will be increased to cover all visible surfaces of the cave as much as possible; moreover, using multi-level shooting, providing multiple high-definition shooting equipment, and ensuring good lighting conditions, each part of the cave will have at least two different shooting angles to obtain a higher quality dataset. Table 1 shows the Comparison between Colmap and OpenMVS operating condition

Table 1.Compare between Colmap and OpenMVS operating condition

Tools	Process	Times	Points
Colmap	Feature extract	0.539m↲	27584 points
	Feature matching	1.705m←	
	Sparse point cloud	1h29m49s	
OpenMVS	Estimated depth-maps	2h02m27s729ms	878190points
	Filtered depth-maps	1m2s415ms	
	Fused depth-maps	658ms	
	ReconstructMesh	3m	
	RefineMesh	2h23m45s	
	TextureMesh	2h56m23s	

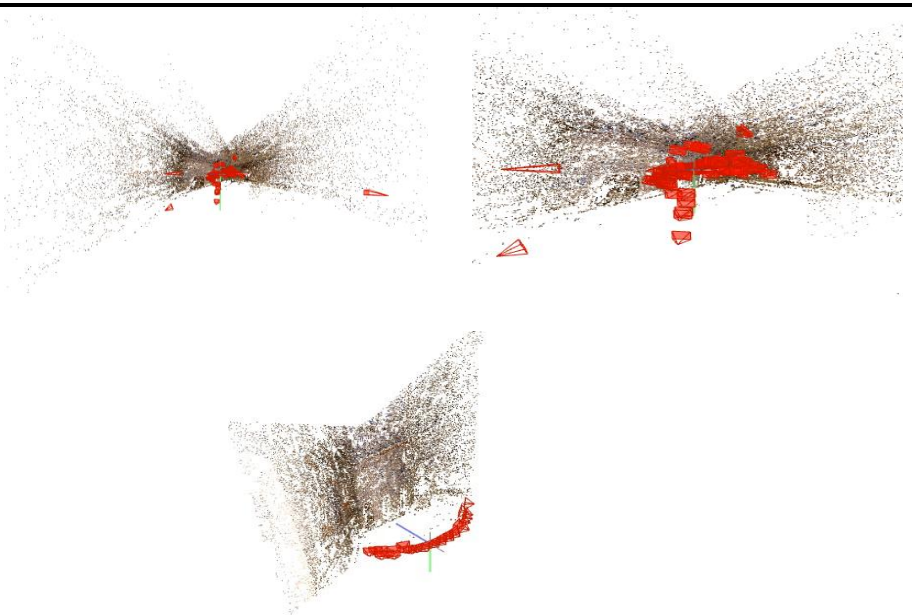

Fig. 1. Diverse angles of the sparse point cloud result

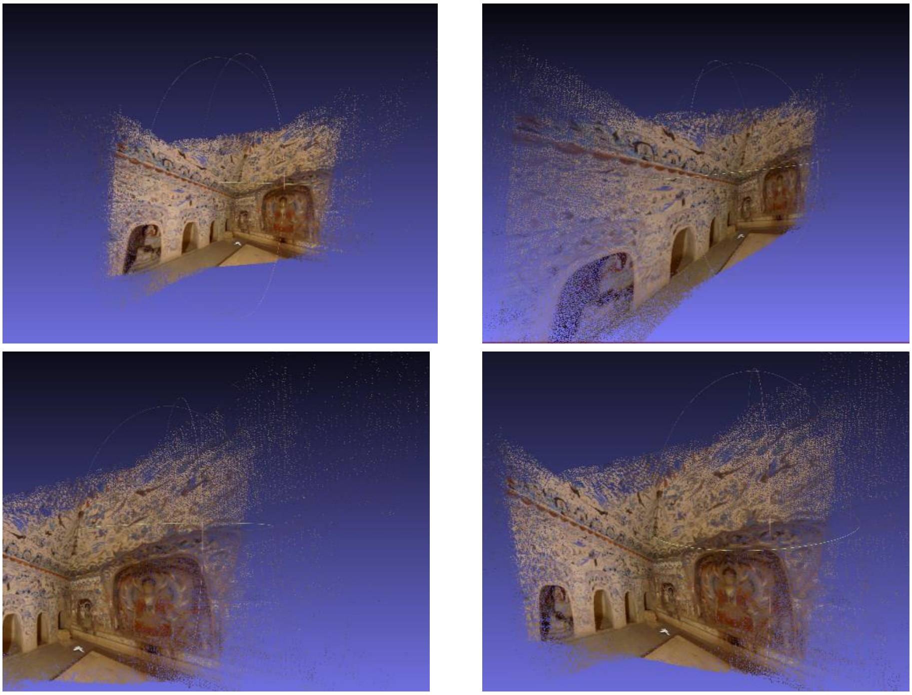

Fig. 2. Diverse angles of the dense point cloud result after texture application

5 Discussion

Based on the analysis and research of MVS algorithms and applications, this technology can perform high-precision modeling of the Mogao Caves artifacts by analyzing data from multiple angles and high resolutions, accurately capturing the detailed features of the artifacts and meticulously recording the damages they have suffered, such as weathering, cracks, and degradation. Thus, this technology is crucial for the restoration and precise repair of cultural heritage. Furthermore, since traditional artifact restoration methods are mostly contact-based inspections and measurements, damages caused by human factors or technological limitations can lead to more severe damage to the artifacts. For instance, the wall paintings in Cave 285 of the Mogao Caves have severely faded due to exposure to light and changes in humidity and air, and the facial contours of the main chamber Buddha sculpture have disappeared. Using MVS for 3D modeling can enhance the safety of the Mogao Caves artifacts and provide a more comprehensive observation of the artifacts. Lastly, the process and results of 3D modeling achieved by MVS technology can facilitate the digitalization of artifacts. Since Cave 285 is one of the earliest excavated caves with recorded dates, over the long span of time, natural disasters, wars, and changes in ancient dynasties have led to the loss of many artifacts. The digitalization of Mogao Caves artifacts can preserve the longevity of artifact information and, at the same time, facilitate the establishment of online exhibitions and digital museums. For example, in recent years, the Dunhuang Research Institute has been collecting data on some of the caves, using information technology to create the "Digital Dunhuang" online platform, which allows for VR presentations of panoramic views of the cave paintings, thus better ensuring the continuation and transmission of culture

The application of MVS algorithms also has several limitations. Firstly, the most critical issue is that MVS technology modeling demands excessive computational resources. As the modeling involves intensive computation, the algorithms themselves are complex and require the processing of vast amounts of data, leading to prolonged execution times and excessive consumption of GPU resources. For large and complex cultural heritage groups like the Mogao Caves, the time required for restoration modeling is too long, making it unsuitable for urgent repair projects. Secondly, the initial collection of artifact data requires specialized technical knowledge, including a thorough understanding of lighting conditions, camera settings, and various shooting and measuring instruments; moreover, the technology is heavily affected by environmental factors. The complex and unstable environment of the Mogao Caves can greatly impact data collection, thereby significantly reducing the quality and accuracy of model reconstruction. Furthermore, MVS technology modeling generates a large amount of high-quality data. To ensure the security and long-term preservation of this data, periodic maintenance and upgrades of storage facilities are necessary, increasing the cost and difficulty of preserving the Mogao Caves.

To address the limitations associated with the technology, several suggestions are proposed to resolve the technical issues. Firstly, with the maturing of cloud computing technology, MVS computational capabilities can be enhanced through cloud services, optimizing algorithm efficiency and facilitating technical and information sharing among cultural heritage institutions. Additionally, using parallel computing and optimizing the algorithms with incremental processing allows for partial data to be processed first, gradually building the model and optimizing the computational resources of MVS technology. Secondly, the development of adaptive lighting condition photography equipment is recommended, using drones or robots for stable shooting in complex environments. This could be complemented with various types of sensors, such as infrared or laser scanning, to capture more details of the artifacts. High Dynamic Range Imaging (HDR) technology can also be utilized to improve issues like overexposure in photographs. Lastly, distributed storage technology should be employed to achieve efficient storage and compression of three-dimensional data, ensuring data security.

Regarding future technological innovations, as deep neural networks mature, convolutional neural networks could be integrated into the feature extraction and matching phases of MVS, enabling self-learning of optimal feature representations. Moreover, based on image segmentation technology, global optimization algorithms could be refined using techniques such as trust region or region-growing, allowing for iterative improvements and gradually integrating smaller model regions into larger structures to optimize overall disparity. Finally, sensor fusion technology should be developed for multimodal data integration, enabling real-time processing of sensor data and continual updates and optimizations of the model.

6 Conclusion

To sum up, MVS 3D reconstruction technology has a great role in promoting the restoration and digital protection of cultural relics in Mogao Grottoes. Compared with other technologies, such as only using image segmentation or machine learning to deduce and restore the damage of murals, this technology collects multi-angle and high-definition data on cultural relics such as murals and sculptures without contact, reduces secondary damage, captures a large number of details and features of the cave, and restores and displays the three-dimensional model of the cave to the greatest extent through continuous optimization and texture mapping of the model, and finally provides features for archaeologists. Information such as color and material can also be simulated and repaired to provide solutions and suggestions for subsequent real restoration work. In future research, we can have an in-depth understanding of the environmental adaptability, multimodal data fusion and efficient computing power of MVS 3D reconstruction technology. With the continuous development of computer vision technology, we will focus on the integration of VR technology and integrate the digital presentation of cultural relics into the concept of the metaverse.

References

1. Fan J.“Management and Monitoring of World Cultural Heritage Sites Based on World Cultural Heritage Values: A Case Study of the Mogao Grottoes of Dunhuang.” Dunhuang Research ,(06):1-5+114(2008).
2. Wang, X., et al. "The history of rescuing reinforcement and the preliminary study of preventive protection system for the cliff of Mogao Grottoes in Dunhuang, China." Heritage Science 9.1: 58 (2021).
3. Fan J. For Dunhuang Grottoes Forever Existence. Dunhuang Research 03:5-9+111(2004).
4. Pan Y., and Lu D.. " Digital Protection and Restoration of Ancient Dunhuang Murals." Journal of System Simulation 15.3: 310-314 (2003).
5. Yu, T., et al. "Artificial intelligence for Dunhuang cultural heritage protection: the project and the dataset." International Journal of Computer Vision 130.11: 2646-2673 (2022).
6. Jin, H., Stefano S., and Anthony J. Yezzi. "Multi-view stereo reconstruction of dense shape and complex appearance." International Journal of Computer Vision 63: 175-189 (2005).
7. Pons, J., Renaud K., and Olivier F.. "Multi-view stereo reconstruction and scene flow estimation with a global image-based matching score." International Journal of Computer Vision 72: 179-193 (2007).
8. Schmid, K., et al. "View planning for multi-view stereo 3D reconstruction using an autonomous multicopter." Journal of Intelligent & Robotic Systems 65: 309-323 (2012).
9. Kuchitsu, N., et al. "Simple evaluation of the degradation state of cultural heritage based on multi-view stereo." Progress in Earth and Planetary Science 6: 1-9 (2019).
10. Hirschmuller, H.. "Stereo processing by semiglobal matching and mutual information." IEEE Transactions on pattern analysis and machine intelligence 30.2: 328-341, (2007).
11. Browning, G., Heinz-Otto K., and Joseph O.r. "Mesh refinement." Mathematics of computation 27.121: 29-39(1973)