Scholar Publishing Group
International Journal of Engineering Technology and Construction
https://doi.org/10.38007/IJETC.2025.060104
ISSN 2789-7176 Vol. 6, Issue 1: 36-43

Design of a 3D Reconstruction System for Immovable Cultural Relics Based on Multi View Stereo Matching and Nerf Fusion

Chuying Lu[1, a*]

[1]*University of Michigan, Ann Arbor 48109, MI, United State*

[a]*chuyinglu2226@gmail.com*

[*]*Corresponding author*

Keywords: Non movable cultural relics, neural radiation field, multi view stereo matching, 3D reconstruction, digital protection

Abstract: The focus of this paper is to develop a 3D reconstruction system for preserving fixed cultural relics, utilizing enhanced neural radiation field (NeRF) and multi view stereo matching technology. This study introduced octree partitioning and dynamic resolution rendering to optimize NeRF visualization and minimize costs. Compared to alternative technologies such as radar and structured light scanning, NeRF is an economically efficient, highly automated, and seamless solution. In addition, Gaussian mapping is used to improve the quality of the reconstructed model. In a case study involving important cultural heritage, NeRF showed better results compared to traditional methods. The system integrates data management, NeRF reconstruction, and visualization functions. However, as the number of partitions increased, it encountered challenges such as compatibility issues and performance degradation. Future research directions will focus on improving the model transformation process, developing efficient partitioning algorithms, and implementing dynamic node addition to enhance system flexibility and stability. These advances aim to support the digital protection of cultural relics and raise public awareness of their importance.

1. Introduction

The design of a 3D reconstruction system for immovable cultural relics based on multi view stereo matching and neural radiation field (NeRF) fusion aims to address the protection challenges faced by China's abundant immovable cultural relics resources. These cultural relics not only carry profound historical and cultural value, but also face multiple threats such as natural disasters, industrial pollution, and insufficient awareness of protection. Digital technology, especially 3D reconstruction technology, provides strong support for the long-term sustainable protection of cultural relics. The 3D reconstruction technology mainly includes methods such as LiDAR, structured light, and visual 3D reconstruction. Although LiDAR has high precision, its cost is high. Structured light technology can achieve millimeter level accuracy but is limited by light and environment, while visual 3D reconstruction technology can achieve large-scale scene 3D reconstruction at a lower cost. Although the accuracy is slightly lower in some cases, the detailed information is rich. In response to the problem of insufficient funds for the protection of low-level and non level immovable cultural relics, visual 3D reconstruction technology has become the preferred choice. Traditional visual 3D reconstruction methods suffer from problems such as model voids and high repair costs. This study proposes a multi view 3D reconstruction method based on NeRF, combined with multi view stereo matching technology, aiming to provide a practical solution with complete models and controllable costs. By integrating the advantages of LiDAR and structured light technology and overcoming their limitations, this system will provide new solutions for the digital protection of immovable cultural relics, promote the development of digital protection of cultural relics, and enhance public awareness and protection of cultural heritage.

2. Correlation Theory

In recent years, a large amount of research has focused on exploring the three-dimensional reconstruction technology of multi perspective images, as well as the application of this technology in various fields such as digital preservation of cultural relics, photogrammetry, and satellite image processing. Researchers have developed a series of innovative methods aimed at improving the accuracy and efficiency of 3D reconstruction. A visibility aware pixel level view selection strategy is proposed to enhance the accuracy of multi view stereo matching; Research has also designed multi view stereo depth inference methods aimed at reducing noise interference, providing more reliable depth information for 3D reconstruction. The efficiency and accuracy of photogrammetric applications have been significantly improved through parallel multi view patch matching technology. In terms of satellite image processing, the construction of a multi view stereo satellite image 3D reconstruction framework based on deep learning provides a new approach for remote sensing data processing. In the application research of neural radiation field technology, researchers have also made significant progress. They demonstrated a multi viewpoint naked eye 3D display technology based on representation fusion NeRF, bringing revolutionary changes to the 3D visual experience. The fusion of cameras and LiDAR provides a new solution for grid assisted neural representation. The interactive fusion technology using neural radiation field has optimized the process of virtual facility inspection, improving inspection efficiency and accuracy. There are also studies that have implemented adaptive sampling refinement in explicit NeRF, further improving the reconstruction quality. In the field of cultural relic protection, a large amount of research has also focused on how to use 3D reconstruction technology to provide support for the digital recording and display of cultural relics. Researchers have achieved 3D reconstruction of movable cultural relics based on significant region optimization techniques, providing strong support for the digital protection of cultural relics. A natural disaster risk monitoring system for immovable cultural relics has been constructed using digital twin technology, providing a scientific basis for cultural relic protection. These studies not only promote the development of 3D reconstruction technology, but also make important contributions to the protection and inheritance of cultural heritage.

3. Method

3.1. Camera Imaging Principle and 3D Reconstruction Technology

The principle of camera imaging is the process of mapping objects in three-dimensional space to a two-dimensional imaging plane, which involves both internal and external parameters of the camera. In the simplified camera pinhole imaging model, light is projected through the camera's pinhole onto the imaging plane to form an image, establishing a mapping relationship between two-dimensional image points and real-world three-dimensional points. In order to clearly illustrate this mapping principle, we have defined four basic coordinate systems: camera coordinate system, world coordinate system, image coordinate system, and pixel coordinate system. They determine the size, shape, and position of objects in the image captured by the camera. The external parameters describe the position and orientation of the camera in three-dimensional space, including the rotation matrix R and the displacement vector .By reverse reasoning, that is, based on the points in the two-dimensional image and the internal and external parameters of the camera, we can reconstruct the points in the three-dimensional scene. In the 3D reconstruction of immovable cultural relics, this process is particularly important, as it enables us to achieve digital protection of cultural relics through multi perspective images and improved neural radiation field (NeRF) technology, providing strong support for the long-term preservation and research of cultural relics.

3.2. NeRF Fundamental Theory and Its 3D Reconstruction Process

NeRF predicts the color and density of voxels in a scene by learning from limited perspective images, and renders to generate views from any perspective. Its input consists of camera centroid coordinates X and ray vector D, and its output is the predicted color C of the view pixel. NeRF's 3D reconstruction includes steps such as voxel sampling, data encoding, neural network training, and image rendering. In NeRF, the MLP (Multi Layer Perception) structure is its core component, including Coarse and Fine, two MLPs with the same DeepSDF structure. Its sampler extracts training data from multi-view images, transformed to a common coordinate system. Positional Mapping in Fourier space helps NeRF fit high-frequency data. Outputs are voxel color and density, visualized by overlaying all voxels a light ray passes through. By using classic rendering equations, the color value of the pixel can be calculated. After rendering voxel samples of each pixel's light direction within the field of view, the final rendered view is formed. During the training process, NeRF dynamically adjusts the internal parameter weights based on the difference between the rendered view and the actual view, ultimately obtaining the NeRF 3D model. By continuously optimizing this loss function, NeRF can gradually improve the accuracy and quality of 3D reconstruction.

3.3. Principles of SFM Algorithm

The SFM algorithm is a core step in multi view 3D reconstruction technology, which provides the basis for reconstructing scene structures for various 3D reconstruction methods including NeRF. This algorithm does not rely on the order of image capture and can restore the three-dimensional structure of objects from multiple unordered images. Its working principle is to estimate the internal and external parameters of the camera when taking each image by analyzing a set of images taken from different perspectives and utilizing the feature point relationships between the images.In SFM, feature extraction uses SIFT to detect stable, rotation, scale, and brightness-invariant features. It involves four stages: scale space extremum detection, keypoint localization, direction assignment, and descriptor generation. Matching uses brute force or KD tree-based search, with KD tree improving efficiency. Matching succeeds if the Euclidean distance between descriptors is below a threshold.The relationship between the intrinsic matrix E and the fundamental matrix F can be expressed by the internal parameter matrix K of the camera. In order to extract the rotation matrix R and the displacement vector t from the intrinsic matrix E, the algorithm performs singular value decomposition (SVD) and selects the correct one from four possible solutions based on the decomposition results and additional scene information or assumptions. By integrating these rotation and translation information as well as the internal parameters of the camera, the SFM algorithm can restore the sparse point cloud structure of the object. This sparse reconstruction process provides important foundational data for subsequent 3D reconstruction tasks. The entire SFM algorithm process not only promotes the development of 3D reconstruction technology, but also provides strong technical support for multiple fields including cultural relic protection.

4. Results and Discussion

4.1. NeRF Reconstruction Optimization, Data Acquisition, and Coding Evaluation

In the field of NeRF-based 3D reconstruction, this research endeavors to enhance model quality and rendering speed, while optimizing the experimental data acquisition process. To improve model quality, a Gaussian Mapping encoding strategy is adopted to replace the original Positional Mapping encoding. By expanding the value range differences in the input data, Gaussian Mapping more effectively captures high-frequency information, significantly boosting NeRF's performance in detail restoration and scene reconstruction. Experiments demonstrate a notable improvement in view quality and detail with this encoding approach.Addressing the slow rendering speed of NeRF models, a dynamic resolution rendering mechanism is introduced. This mechanism allows for dynamic adjustment of view resolution during rendering, reducing the GPU workload and maintaining stable frame rates, thereby significantly improving rendering speed. For experimental data acquisition, a DJI Matrice 350 RTK drone equipped with a Zenmuse P1 camera is used to capture high-resolution and high-precision images. To minimize interference from dynamic objects in the data, six specific time periods were chosen, and within each, images with minimal interference were selected to form the final dataset for modeling. To enhance experimental testing efficiency, ultra-high-resolution images were processed while retaining critical metadata. Subsequently, Colmap software was used for SFM sparse reconstruction, yielding a sparse 3D model and camera intrinsic and extrinsic parameters, which provided essential input for subsequent NeRF-based 3D reconstruction.To evaluate the effectiveness of Positional Mapping and Gaussian Mapping encoding strategies, the experimental environment was configured with an Intel i9 13900k processor, NVIDIA RTX 4090 GPU, running Ubuntu 22.04, CUDA 11.8.0, and Python 3.8.0. The NeRF model, developed based on the tiny-cuda-nn framework, was set with both uniform sampling points and probabilistic density sampling points at 64, and underwent 15 rounds of training. A simplified four-layer network structure, with 256 neurons per layer, was constructed for the image regression task to mimic the scene regression task in NeRF-based 3D reconstruction. The result data is shown in Figure 1

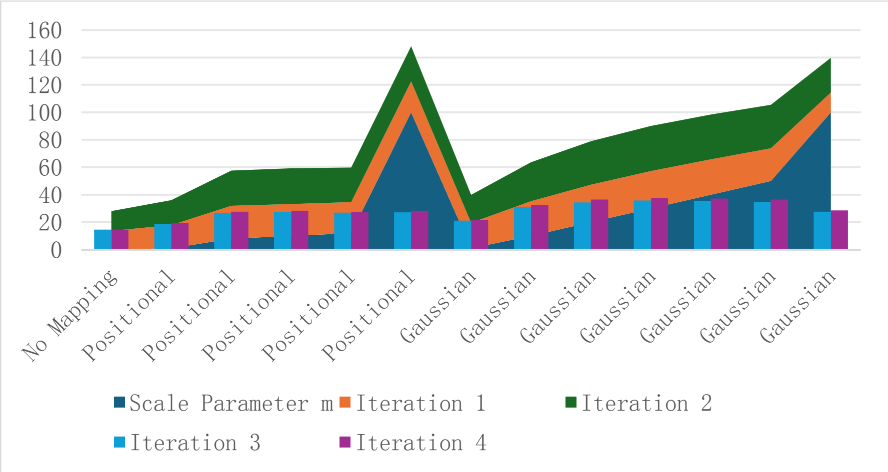

Figure 1. PSNR Comparison Table

The experimental results indicate that encoding the input data can substantially elevate the PSNR value in MLP image regression tasks. However, the choice of the scale parameter m must be approached with caution, as values that are either too low or too high may induce premature convergence in the model, thereby constraining its learning capacity. These discoveries hold significant importance for refining the input data encoding strategy in NeRF-based 3D reconstruction. This chapter delves into the process of 3D reconstruction for immovable cultural relics and compares the efficacy of Metashape, NeRF, and MVS technologies.In the application of Metashape, the procedure begins by uniformly formatting POS data. Depending on the size of the modeling area, different strategies are employed to import image sets. Photogrammetric principles are then utilized to extract point clouds, which are utilized to construct a 3D model. Materials are assigned to the model, and ultimately, precise digital surface models and digital orthophotos are output. The results are shown in Table 1

Table 1. Evaluation of 3D Reconstruction Quality Using Different Methods

View/Method	NeRF	MVS	Metashape
Top-view PSNR↑	23.03	23.02	28.14
Front-view PSNR↑	25.53	22.65	26.35
Left-view PSNR↑	25.36	23.89	27.37
Eaves PSNR↑	24.84	18.78	15.61
Top-view SSIM↑	0.55	0.53	0.75
Front-view SSIM↑	0.68	0.42	0.68
Left-view SSIM↑	0.62	0.52	0.72
Eaves SSIM↑	0.69	0.43	0.21
Top-view LPIPS↓	0.52	0.54	0.32
Front-view LPIPS↓	0.45	0.59	0.49
Left-view LPIPS↓	0.48	0.51	0.40
Eaves LPIPS↓	0.33	0.62	0.79

In contrast, NeRF demonstrates remarkable performance in terms of model stability and accuracy, with a distortion rate lower than MVS and approaching that of Metashape. Additionally, NeRF outperforms MVS in three image quality evaluation indicators: PSNR, SSIM, and LPIPS, with an overall performance nearly on par with Metashape. Notably, when utilizing the Gaussian

Mapping data encoding strategy, the PSNR value reaches 37.51, substantially higher than the 28.28 achieved with Positional Mapping. To address NeRF's slower rendering speed, this chapter examines dynamic resolution technology, which enhances the user experience by filling the waiting time with incremental rendering.The optimization of the Gaussian Mapping data encoding strategy and the implementation of dynamic resolution technology have significantly improved the applicability of NeRF in the field of immovable cultural heritage digitization, enabling more refined views and real-time visualization.

4.2. Improvement of NeRF System Based on Octree and Distributed Training

This study proposes a method combining octree optimization and distributed training system for large-scale 3D reconstruction of immovable cultural relics. Firstly, the scene size and point cloud are obtained through sparse reconstruction, and the scene is segmented using an octree. Then, voxels are merged into the target partition through breadth first search. Each partition contains several voxels, and the image containing the partition scene is found, and the mapping information between the image and the partition is recorded. Subsequently, the images of each partition are input into NeRF training to obtain the region model, which is finally logically merged into a complete 3D model. This improvement allows for only replacing defective partitions and re rendering, saving time and resources.

In terms of samplers, linear differential samplers are used to improve modeling accuracy. Meanwhile, design a distributed training framework based on Hadoop, including storage and computing components. The storage component adopts a master-slave node architecture to manage model naming, mapping, and storage tasks. The computing component is responsible for modeling and visualizing the regional model, which is divided into two stages: distribution and merging. Inter node communication relies on Socket technology, with the master node responsible for task allocation and tracking, and the slave nodes responsible for training regional models. During rendering, the master node determines the partition through which the light passes, assigns voxel sampling tasks to the slave nodes, and finally aggregates and calculates pixel colors.Cluster environment configuration includes connecting to the local area network, assigning static IP addresses and gateways, setting unique host names and IP mappings, opening necessary ports, creating dedicated user accounts and configuring password free login, installing Python interpreter and adding environment variables. When installing a distributed training system, import the NeRF code into the master node, modify the configuration file, and use the scp command to copy it to the slave nodes for unified installation. This improved NeRF system provides an efficient and flexible solution for large-scale 3D reconstruction.

4.3. Comparative Analysis of Evaluation Effects

In this chapter, we investigated the challenges of NeRF in large-scale immovable cultural relic scenes, including high computational requirements and local unrepairability. The scene segmentation based on octree effectively solves these problems and enhances the flexibility and practicality of NeRF. It reduces complexity and computational requirements, especially in resource limited environments. Distributed training can improve efficiency in resource rich environments. Our improved segmentation significantly enhances the applicability of NeRF in large-scale scenarios., We set 64 uniform and probabilistic sampling points for NeRF and trained on an 8-machine cluster with Intel i9 13900k and NVIDIA RTX4090. We test segmentation by dividing sparse point clouds into voxels and merging them into different partitions, recording the average time and memory for each partition. The results show that the training data for each partition decreases as the number of partitions increases, but this relationship is not strictly proportional due to image overlap. A reasonable number of partitions can optimize performance. For Tianhou Palace, four partitions are the most cost-effective. To verify the accuracy of linear interpolation under uniform sampling, we model NeRF using both as MLP inputs. The PSNR results show that the view can be improved through linear interpolation.

5. Conclusion

This article investigates an improved NeRF based 3D reconstruction of immovable cultural relics and proposes octree segmentation and dynamic resolution rendering to address the challenges. Compared to radar and structured light scanning, NeRF provides a cost-effective, highly automated, and pore free model. Gaussian mapping improves view quality. In the case of Tianhou Palace, compared with MVS and Metashape, NeRF reduced distortion by 2.32% and 0.08% respectively, and performed well in terms of image quality indicators. Dynamic resolution rendering optimizes the visualization effect. Octree partitioning reduces coupling, enhances local replacement and repair, and improves training efficiency through a distributed framework. The resulting system integrates data management, improved NeRF reconstruction, and dynamic visualization, reducing technical barriers. The NeRF model is framework specific and incompatible with traditional software, and its performance decreases with increasing partitioning. Future work will explore model transformation, efficient partitioning, and dynamic node addition to enhance flexibility, stability, and support for large-scale reconstruction.

References

[1] Xu, Yue. "Research on Maiustream Web Database Development Technclogy." Journal of Computer Science and Artificial Intelligence 2.2 (2025): 29-32.

[2] Zhu, Zhongqi. "Strategies for Improving Vector Database Performance through Algorithm Optimization." Scientific Journal of Technology 7.2 (2025): 138-144.

[3] Wang, Buqin. "Strategies and Practices for Load Test Optimization in Distributed Systems." Scientific Journal of Technology 7.2 (2025): 132-137.

[4] Wu, Linwei. "Research on Data Integration and Process Optimization in the Field of Financial Technology." International Journal of Finance and Investment 2.1 (2025): 82-86.

[5] Zhang, Jingtian. "Research on Worker Allocation Optimization Based on Real-Time Data in Cloud Computing." Frontiers in Science and Engineering 5.2 (2025): 119-125.

[6] Hao, Linfeng. "Application of Machine Learning Algorithms in Improving the Performance of Autonomous Vehicles." Scientific Journal of Technology 7.2 (2025): 118-124.

[7] Pan, Yu. "Research on the Evolutionary Path of Resource Management and Capability Building for Platform Enterprises." International Journal of Finance and Investment 2.1 (2025): 78-81.

*[8] Gu, Yiting. "Practical Approaches to Develo**High-performance Web Applications Based on React." Frontiers in Science and Engineering 5.2 (2025): 99-105.*

[9] Cui, Naizhong. "Optimization Strategies for Traffic Signal and Identification Design." Frontiers in Science and Engineering 5.2 (2025): 92-98.

[10] Li, Xuan. "Research on the Review Technology of Building Fire Protection Design Drawings Based on BlM." Frontiers in Science and Engineering 5.2 (2025): 113-118.

[11] Xu Y. Research on UAV Navigation System Based on Behavioral Programming[C]//2024 IEEE 7th International Conference on Automation, Electronics and Electrical Engineering (AUTEEE). IEEE, 2024: 419-425.

[12] Chen, H., Yang, Y., & Shao, C. (2021). Multi-task learning for data-efficient spatiotemporal modeling of tool surface progression in ultrasonic metal welding. Journal of Manufacturing Systems, 58, 306-315.

[13] Ma, K., Zhang, N., Mei, X., Feng, C., Hou, W., & Ye, Z. (2024, October). Research on Optimization of Shared Bicycle Scheduling Based on Genetic Algorithm and LSTM. In 2024 IEEE 6th International Conference on Civil Aviation Safety and Information Technology (ICCASIT) (pp. 936-940). IEEE.

[14] Xiang, Y., Li, J., & Ma, K. (2024, October). Stock Price Prediction with Bert-BiLSTM Fusion Model in Bimodal Mode. In Proceeding of the 2024 5th International Conference on Computer Science and Management Technology (pp. 1219-1223).

[15] Tan, Weiyan, Shujia Wu, and Ke Ma. "Freight Volume Prediction for Logistics Sorting Centers Using an Integrated GCN-BiLSTM-Transformer Model." Advances in Computer and Engineering Technology Research 1.4 (2024): 320-324

[16] Fan, Sunjia, et al. "Defense methods against multi-language and multi-intent LLM attacks." International Conference on Algorithms, High Performance Computing, and Artificial Intelligence (AHPCAI 2024). Vol. 13403. SPIE, 2024.

[17] Ma Z. Strategies for Enhancing Customer Lifetime Value through Data Modeling[J]. European Journal of Business, Economics & Management, 2025, 1(1): 1-7.

[18] Li, X. (2025). Research on Three - dimensional Modeling of Urban Buildings based on CityGM. Scientific Journal of Technology, 7(3), 302-306

[19] Wang, Yuxin "Research on Intelligent Macro Image Recognition Algorithm of Oil Pipe Failure Based on Deep Learning." Journal of Image Processing Theory and Applications (2025), 8(1): 1-7

[20] Zhang Y. Research on Optimization of Engineering Cost Database Based on Big Data and Intelligent Technology[J]. International Journal of New Developments in Engineering and Society, 2024, 8(5).

[21] Li, X.(2025)"Research on the application of GPS, total station and CAD Technology in architectural Grid." Computer Life (2024),12(3),36-39.

[22] Zhang, Jinshuo "Research on Real Time Condition Monitoring and Fault Warning System for Construction Machinery under Multi Source Heterogeneous Data Fusion." Journal of Engineering Mechanics and Machinery (2024), 9(2): 139-144

[23] Zhang, Yiru. "Design and Implementation of a Computer Network Log Analysis System Based on Big Data Analytics." Advances in Computer, Signals and Systems,(2024) 8(6),40-46.

[24] Xu, Y. (2025). Research on Maiustream Web Database Development Technclogy. Journal of Computer Science and Artificial Intelligence, 2(2),29-32

[25] Yang, Jinzhu "Integrated Application of LLM Model and Knowledge Graph in Medical Text Mining and Knowledge Extraction."Social Medicine and Health Management (2024), 5(2): 56-62

[26] Fan, Yijiao "Research on Risk Spillover Measurement of Fintech System Based on DCC-GARCH and Generalized Variance Decomposition Network Model."Accounting, Financial Engineering and Risk Management (2024), 7(5): 43-50