3D Reconstruction of Indoor Scenes Based on 3DGS Models

Hanghua Li
School of Computer Science and Engineering
Xi'an Technological University
Xi 'an, China
E-mail: 2451734651@qq.com

Lipeng Si
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: 37648537@qq.com

***Abstract*—With the rapid development of computer vision and artificial intelligence technologies, indoor scene reconstruction has been more and more widely used in the fields of virtual reality, augmented reality and architectural design. In this paper, we study an indoor scene reconstruction method based on the 3DGS model, which has been widely used in computer graphics and vision processing with powerful scene representation and rendering capabilities. In this study, we optimize the 3DGS model to enhance the detail preservation and realism of the reconstruction results by adjusting the opacity of the Gaussian function. We used the Replica dataset and the self-harvested dataset for model training. Through experimental validation, the peak signal-to-noise ratio as well as the structural similarity ratio of the reconstruction results of the optimized model have an improvement effect of more than 1%, which indicates that the optimized model has a significant improvement in detail retention and realism, and the reconstructed scene performs more realistically in terms of texture details and light and shadow effects.**

Keywords-3DGS; Indoor Scene; 3D Reconstruction

I. INTRODUCTION

In recent years, there are more and more demands for indoor refined 3D models in smart cities, cultural relics protection, indoor navigation, virtual reality, etc. 3D reconstruction of indoor scenes has become one of the important research topics in the field of computer vision and computer graphics. Scene 3D reconstruction refers to the acquisition of image or video data of the indoor environment, the use of computer vision technology and 3D reconstruction algorithms to analyze the image content and geometric information, to infer the layout of the room, the position and size of the furniture, the geometry of the walls and floors and other structured information, and ultimately to construct a real indoor 3D model.

Traditional scene reconstruction methods rely on camera position and pose information as well as data from depth sensors, but these methods usually suffer from limitations in real-time, accuracy, and cost. With the advent of the 3DGS algorithm, it enables high-quality real-time rendering and scene optimization by efficiently modeling the scene using Gaussian functions. The technique starts from a sparse point cloud, represents the scene as a differentiable 3D Gaussian set, and constructs an accurate and compact representation of the scene by optimizing its properties such as position, opacity, and covariance. During the rendering process, the 3DGS algorithm utilizes fast GPU sorting and tile-based rasterization to achieve efficient visibility-aware rendering with anisotropic splash support, thus achieving real-time rendering while ensuring rendering quality.

In this paper, we will use 3DGS based algorithm to reconstruct the indoor scene, which can reconstruct the complex indoor scene efficiently and reliably. And the model is trained on Replica dataset as well as self-collected dataset and further optimized. The experimental results show that the performance of the model is improved.

II. RESEARCH BACKGROUND

In recent years, 3D scene reconstruction technology has made significant progress in the fields of computer graphics and computer vision, which refers to the transformation of two-

dimensional image or video data into interactive 3D models through computer vision, 3D reconstruction, deep learning and other technical means. This process involves multiple technical fields, including computer graphics, image processing, machine learning, etc. and aims to accurately restore the physical space and generate 3D digital models with a high degree of realism.

The 3D scene reconstruction technique mainly includes several steps of data acquisition and processing, feature extraction, 3D reconstruction, and result optimisation. Among them, the dataset is mainly categorized into point cloud, mesh and voxel, and the 3D reconstruction mainly includes indoor scene reconstruction based on deep learning and 3D reconstruction by traditional methods. PointNet [1], a deep learning network that directly processes point cloud data, proposed by researchers at Stanford University, provides an effective solution for 3D reconstruction of indoor scenes. NeRF Neural Radiation Field, an emerging technique proposed by Ben Mildenhall et al [2], utilizes a neural network model to achieve fast reconstruction and rendering of indoor scenes by training on captured scene images to learn attributes such as lighting, material, and depth of the scene, and then generating realistic images of new perspectives. Neural radiation fields have become an important area of research in subsequent research efforts. The main ones include D-NeRF [3], which is able to learn dynamic deformable fields from image view sequences, NSFF [4], a scene flow field algorithm for the free synthesis of spatio-temporal views of dynamic scenes, NeRV [5], a neural reflective and visible field algorithm for view and illumination resynthesis, and GIRAFFE [6], a composable generative feature algorithm for editable scene representations. While all of the above techniques can be reconstructed for different scenes, the aforementioned neural radiation field models are mainly deep convolutional neural networks, which take much longer to train compared to traditional shader and illumination techniques, possibly several times longer than these techniques. It also requires significant computational resources to support its training and rendering process.

Recently, a 3DGS technique based on neural radiation field was proposed by Bernhard Kerbl et al [7], which has attracted much attention due to its high efficiency and real-time performance. This technique realizes efficient reconstruction and real-time rendering of the scene by utilizing the Gaussian function to represent the spatially continuous distribution of the data, which provides a new way of thinking for the reconstruction of the indoor scene. The 3DGS technique has already shown its advantages in accurate modeling and detail 3DGS technology has demonstrated its advantages in accurate modeling and detail preservation. It not only preserves the geometric information of the scene, but also retains the rich texture and lighting effects during the rendering process.

In conclusion, with the continuous development of research and technology, the advantages of 3DGS algorithm in terms of accuracy, real-time and efficiency will become more and more significant. These advantages will not only improve the quality of indoor scene reconstruction, but also bring more possibilities for interior design, virtual reality, augmented reality and many other fields. At the same time, the continuous progress of 3DGS will not only change the way we understand and utilise interior space, but also provide a broad stage for future technological applications and innovations.

III. Indoor Scene 3D Reconstruction

Indoor scene reconstruction using 3DGS mainly includes the following steps: data acquisition and processing, indoor scene reconstruction and model optimisation. Data acquisition is to collect data from the scene to be reconstructed by using the shooting tool, and then the collected data are preprocessed to be converted into SfM point cloud data, and the scene reconstruction results are obtained by training the 3DGS model. The process of reconstruction is shown in Fig. 1.

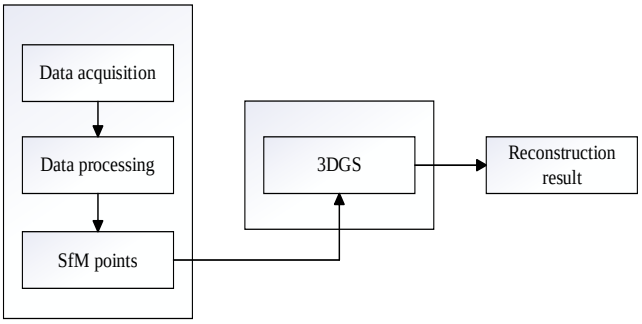

Figure 1. 3D Reconstruction Process Map.

A. Data Acquisition And Processing

Considering the scene equipment variability of the shot, after converting the shot into a continuous image, the denoising process is performed on this continuous set of images. Common image noise processing schemes mainly include mean removal filter for mild noise, Gaussian filter for Gaussian noise, and bilateral filter for filtering noise while preserving image edge information. Therefore, we introduce bilateral filtering to preprocess the data, and realize the removal of noise and smoothing of local edges on the basis of retaining regional information by comprehensively considering the spatial information of the image within the filter and the similarity of pixel gray values， some of the detail processing results are shown in Figure 2.

Figure 2. The result of adding bilateral filters for indoor scenes.

B. Model Introduction

3DGS modeling is an innovative scene reconstruction and rendering technique based on Gaussian distribution. Its core idea is to use 3D Gaussian distribution as the basic element to represent the geometric and color information in the scene, and to render this information onto a 2D plane by rasterization showing unique advantages in many application scenarios. It mainly includes the following steps:

1) Creating Gaussian Functions

A 3D Gaussian was chosen for this experiment, which can be easily projected into a 2D image, thus allowing for fast blending rendering. Starting with a set of sparse point clouds, each feature point is represented in 3D space by a Gaussian function. The Gaussian function is defined by several parameters, including position, covariance matrix, color, and transparency. Our Gaussian is defined by the full 3D covariance matrix Σ defined in the world space, centered at the mean point:

$$G(x) = e^{(-\frac{1}{2})x^T \sum^{-1}(x)} \tag{1}$$

At the same time, an affine transformation is needed to project the 3D Gaussian to 2D for rendering, letting the covariance matrix in the camera coordinate system be Σ. Given a scaling matrix S and a rotation matrix R, we can find the corresponding Σ:

$$\sum = RSR^T S^T \tag{2}$$

2) Adaptive density control

Starting from the initial sparse point of the SfM, the initially sparse set of Gaussians is changed into a denser set of Gaussians by controlling the number and density of Gaussians per unit volume to better represent the scene. Gaussian densification is performed every 100 iterations and removes essentially transparent Gaussians, i.e., Gaussians with α less than a threshold. Our adaptive control part of the Gaussian needs to be corrected by moving the Gaussian for regions with missing geometric features and regions where the Gaussian covers a large area of the scene. For small Gaussian areas with missing geometric features, they need to be covered. For large Gaussians that are large for the Gaussian coverage of the scene need to be split into smaller Gaussians, using two new Gaussians to replace these Gaussians. In the first case, the need to increase the total volume and the number of Gaussians is detected and handled, and in the second case, the larger Gaussians are split into multiple smaller Gaussians.

3) Rasterization

The screen is first partitioned into 16*16 blocks and the Gaussians with 99% confidence intervals

that intersect the view vertebrae are retained. At the same time, each Gaussian is instantiated according to the number of tiles they cover, and then each instance is assigned a key that combines the depth of the view space and the tile ID. The Gaussians are then sorted based on those keys.

After sorting the Gaussian, the entries are sorted by recognising the first and last depth of the splat to which they were sputtered, and a list is generated for each tile. Rasterisation is then performed, and for each tile a thread block is started, each of which loads the packet containing the Gaussian into shared memory and traverses the list from front to back through the colour and opacity values based on the given pixels, thus simultaneously loading and processing the data. When we reach the target saturation level during the accumulation of pixels in the continuous traversal, the corresponding thread stops, and the processing of the whole tile terminates when the opacity of all pixels is 1.

C. Extracting 3DGS Model Optimization

Since the SfM point cloud data generated by colmap increases to millions when the dataset is large, these data take relatively long time in scene training, which affects the efficiency of 3DGS splatting. In this experiment, a pruning operation on the Gaussian volume is added in the middle, with the main purpose of removing the high-density and high-volume regions in the scene due to false positives or redundancy, in order to improve the rendering quality and reduce the amount of computation. The specific flow of this experiment is shown in Fig. 3.

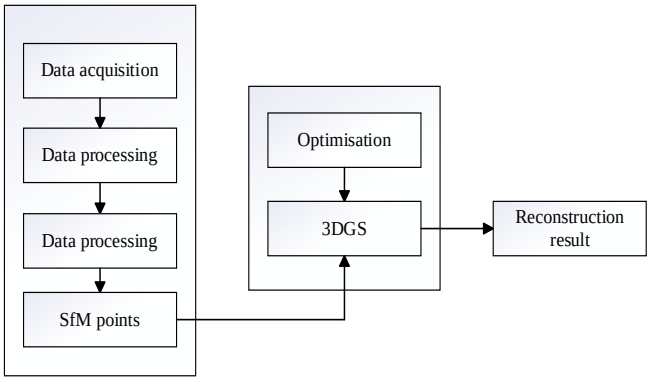

Figure 3. Optimized 3D Reconstruction Process Map.

First, a series of 3D Gaussian bodies are defined within the view body, which are rendered by projecting them to the camera viewpoint so that their impact can be observed in the 2D image plane. For each Gaussian body, we calculate its contribution to each pixel or ray. This is done by determining whether the Gaussian body intersects a light ray. Iterating over all the pixels in the training view, the number of 'hits' on a pixel is calculated for each Gaussian as an initial significance score. In addition to the basic number of 'hits' on a pixel, we also consider the volume and opacity of the Gaussian to further refine the score. So the summary is as follows:

$$GS_j = \sum_{i=1}^{MHW} L(G(X_j), r_j) \cdot \sigma_j \cdot \gamma(\sum\nolimits_j) \quad (3)$$

Where j is the Gaussian index, i is the pixel, and M, H, and W are the number of training views, image height, and image width, respectively. l is the indicator function, which determines whether or not the Gaussian function intersects a given ray. However, the use of Gaussian volume tends to exaggerate the importance of the background Gaussian distribution, leading to excessive pruning of the Gaussian distribution for complex geometric models. Therefore, we introduce a more adaptive method to measure the size of its volume.

$$\gamma(\sum) = (V_{norm})^{\beta} \quad (4)$$

$$V_{norm} = \min(\max(\frac{V(\sum)}{V_{\max 90}}, 0), 1) \quad (5)$$

The range of Gaussian volumes was limited to 0 to 1 by sorting all Gaussian volumes and normalizing the top 90% of maximum values in the Gaussian volume as a benchmark, thus avoiding overly high or underly high floating-point Gaussian values obtained directly from the original 3DGS.

IV. Experimental Analysis

In this paper, we firstly reconstructed the indoor scene using the base 3DGS model, and partially optimised the base 3DGS model, and evaluated the reconstruction quality of this paper's model in the Replica dataset and the self-picked data scene, and quantitatively analysed the reconstruction results by two metrics: PSNR and SSIM.

This paper first uses the original model to reconstruct indoor scene scenes for the Replica dataset, which is rich in scene details with dense meshes, high dynamic range textures, semantic layers, and reflective properties, and record the training results, and then we use the optimized model in this paper to reconstruct each scene in this dataset and record the results to compare with the former results.

The results of comparing the peak signal-to-noise ratio and structural similarity obtained by training the Replica dataset after optimizing the opacity of the original 3DGS model are shown in Table I.

TABLE I. Experimental Results

Dataset	PSNR		SSIM	
	3DGS	ours	3DGS	ours
office0	38.45	39.05	0.925	0.934
office1	36.97	37.17	0.938	0.954
office2	36.39	36.68	0.945	0.962
office3	35.85	36.35	0.941	0.958
office4	36.54	36.66	0.939	0.941
room0	37.26	38.06	0.915	0.923
room1	36.28	37.75	0.929	0.935
room2	34.98	35.14	0.913	0.927

By reconstructing the Replica dataset with diversity, we observe that the optimised method performs well on image quality assessment metrics. Both the Peak Signal-to-Noise Ratio (PSNR) and the Structural Similarity Index (SSIM) show significant performance improvement in all test scenarios. This consistent improvement not only reflects the robustness of the optimised algorithm in dealing with different types of indoor environments, but also indicates its good adaptability in scene reconstruction.

The peak signal-to-noise ratio results obtained by training the Replica dataset after optimizing the opacity of the original 3DGS model are shown in Fig. 4.

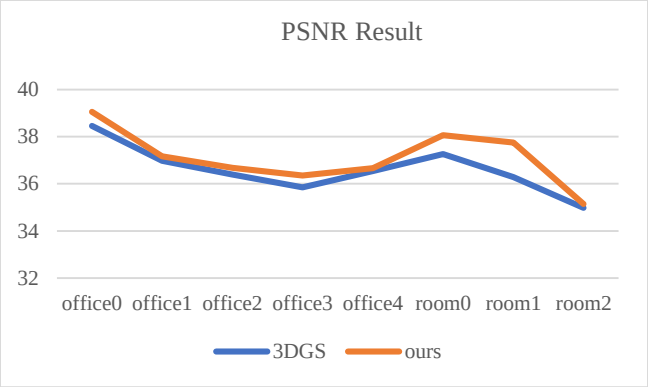

Figure 4. Performance of peak signal-to-noise ratios of 3DGS and ours models for eight different indoor scenes, respectively.

The structural similarity ratio results obtained by optimizing the opacity of the original 3DGS model after training on the Replica dataset are shown in Fig. 5.

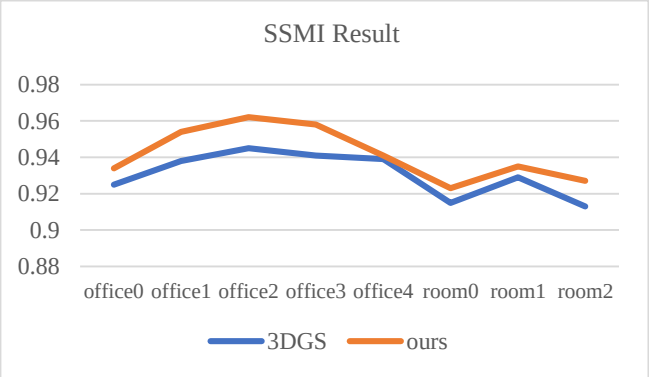

Figure 5. Representation of structural similarity between 3DGS and ours model for 8 different indoor scenes, respectively.

In multiple scenes of Replica dataset, the reconstruction results of the optimized 3DGS model show better results than the original 3DGS model in both PSNR and SSMI. With the increase of the number of Gaussian primitive, the PSNR value shows an obvious upward trend, which verifies the positive correlation between the model scale and the reconstruction quality. Meanwhile, the reconstruction results of the 3DGS model on the Replica dataset are not only highly similar to the real scene in terms of brightness and contrast, but also maintain good consistency in structural information. This indicates that the model is able to accurately capture the structural features of the scene, thus generating visually more realistic reconstruction results.

A. Indoor scene reconstruction on a self-built dataset

In order to evaluate the applicability of the model in this paper more comprehensively, a series of own environmental data were also collected in this study. It is used to verify the performance of the optimized 3DGS model in the paper in real scenarios and to compare the results, as shown in the following table II.

TABLE II. Experimental Results

Dataset	PSNR	SSIM
3DGS	30.41	0.879
ours	31.89	0.897

In the self-collected data scenes, the optimized 3DGS model also shows superior reconstruction quality to the original 3DGS model. Especially in the area with complex light and rich texture, the model can better restore the scene details, and the PSNR value is kept at a high level. the SSIM value is also kept at a high level, which better adapts to the scene changes and maintains the structural similarity of the reconstruction results.

Through this experiment, we verified the excellent reconstruction performance of the model in the Replica dataset and self-collected data scenes, and both PSNR and SSIM metrics show that the model can efficiently and accurately reconstruct the 3D scene and retain rich structural information and visual details. In the future, we can further optimize the model structure and training algorithm to improve its reconstruction efficiency and quality in large-scale and complex scenes.

V. CONCLUSIONS

In order to solve the traditional neural radiation field for indoor scene reconstruction problem, and through the literature research is the scene of the method, the final choice of 3DGS model, selected Replica dataset and self-collected data scene reconstruction quality In the future, we can further optimise the model structure and training algorithms, in order to improve its reconstruction in large-scale, complex scenes in the efficiency and quality, for virtual reality, game development, film production and other fields to provide more possibilities.

REFERENCES

[1] Chen C , Fragonara L Z , Tsourdos A .GAPointNet: Gra ph Attention based Point Neural Network for Exploiting Local Feature of Point Cloud[J].Neurocomputing, 2021, 438(7553).

[2] Mildenhall B , Srinivasan P P , Tancik M ,et al.NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis[C]//2020.

[3] Pumarola A, Corona E, Pons-Moll G and Moreno-Noguer F. D-NeRF: Neural Radiance Fields for Dynamic Scenes [C]. IEEE/CVF Conferenceon Computer Vision and Pattern Recognition (CVPR), 2021:10313-10322.

[4] Li Z, Niklaus S, Snavely N and Wang O. Neural Scene Flow Fields for Space-Time View Synthesis of Dynamic Scenes [C]. IEEE/CVF Conference on Computer Vision andPattern Recognition (CVPR), 2021:6494-6504.

[5] Srinivasan P, Deng B, Zhang X, Tancik M, Mildenhall B and Barron J. NeRV: Neural Reflectance and Visibility Fields for Relighting and View Synthesis [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021:7491-750.

[6] Niemeyer M and Geiger A. GIRAFFE: Representing Scenes as Compositional Generative Neural Feature Fields [C]. IEEE/CVF Conference on Computer Vision and Pat- tern Recognition (CVPR), 2021:11448-11459.

[7] Niemeyer M and Geiger A. GIRAFFE: Representing Scenes as Compositional Generative Neural Feature Fields [C]. IEEE/CVF Conference on Computer Vision and Pat- tern Recognition (CVPR), 2021:11448-11459.