Proceedings of the 2025 SICE Festival with Annual Conference
September 9-12, 2025, Chiang Mai, Thailand

Challenges and Trade-Offs in 3D Reconstructing Degraded Video Data from Nuclear Reactor Environments

Stephanie Nix[1†] and Hirokazu Madokoro[2]

[1]Department of Software and Information Science, Iwate Prefectural University, Takizawa, Japan
(Tel: +81-19-694-2500; E-mail: nix_s@iwate-pu.ac.jp)
[2]Department of Software and Information Science, Iwate Prefectural University, Takizawa, Japan
(Tel: +81-19-694-2500; E-mail: hirokazu_m@iwate-pu.ac.jp)

Abstract: This paper addresses challenges in 3D reconstruction from degraded video data in nuclear reactor environments, focusing on post-Fukushima Daiichi decommissioning tasks. High radiation levels and image degradation due to white noise and particulate matter are critical obstacles for remote inspection using ROVs. To evaluate robust reconstruction methods, the study utilizes video footage from the ROV-A2 used by the Tokyo Electric Power Company, with preprocessing involving frame extraction, text removal, and COLMAP-based parameter estimation for SeaThru-NeRF and 3D-GS. Results comparing these approaches under degraded conditions highlight the advantage DUSt3R has in reconstructing structures of interest with high perceptual realism. Metrics such as PSNR, SSIM, and LPIPS quantify accuracy in reproducing detailed texture visuals from set positions, on which the point-based 3D Gaussian splatting scores highest. These results highlight the trade-off between perceptual realism, favoring continuous volumetric reconstructions, and metric accuracy, favoring discrete point cloud representations with explicit texture encoding.

Keywords: 3D reconstruction, SeaThru-NeRF, 3D Gaussian Splatting, DUSt3R, Nuclear Decommissioning

1. INTRODUCTION

Currently, decommissioning efforts are underway at the Fukushima Daiichi Nuclear Power Station, with a focus on remotely inspecting and removing fuel debris using a variety of ROVs (Remote-Operated Vehicles). However, radiation levels near the ROV entry ports are extremely high, limiting operational duration to short shifts. Additionally, captured video footage is compromised by radiation-induced white noise and reduced visibility due to suspended particulate matter.

To address these issues, this research is part of a broader initiative with the goal to develop 3D reconstruction methods that are robust in areas such as the degradation in image quality and the scattering medium that permeates the environment, in order to assist in decommissioning tasks in melted-down nuclear facilities. This research underscores the critical need for robust, task-specific 3D reconstruction methods tailored to extreme enviroments such as high-radiation underwater environments, with implications for decommissioning strategies in nuclear facilities worldwide.

2. METHOD

2.1. Overview

Current predominant methods in 3D reconstruction include those derived from NeRF (Neural Radiance Fields), Gaussian splatting, and 3D foundation models based on DUSt3R (Dense and Unconstrained Stereo 3D Reconstruction). This study compares the NeRF-derived method SeaThru-NeRF [2], which explicitly considers the medium of filming, with the Gaussian splatting vanilla model, 3D-GS (3D Gaussian Splatting) [3], and the vanilla DUSt3R [6] model.

† Stephanie Nix is the presenter of this paper.

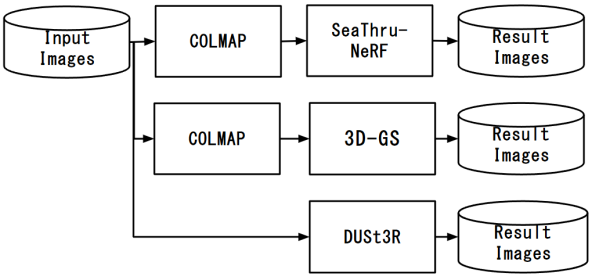

Fig. 1 Overview of the workflow performed in this study.

Both NeRF and 3D-GS require prior camera pose estimation with SfM (Structure from Motion), using implementations such as COLMAP [4]. However, the quality of this estimation depends heavily on the input data, posing a challenge if inaccuracies in pre-processing lead to a failed 3D reconstruction [5]. To address this, we included DUSt3R, an end-to-end 3D reconstruction method that performs camera pose estimation and 3D reconstruction simultaneously, and conducted comparative experiments using the same set of reactor images. This workflow is summarized in Fig. 1.

To visualize the estimated positions of each image and the results of 3D reconstruction, we used Nerfstudio [7] installed on a compute node with 2 NVIDIA RTX 4000 Ada GPUs.

2.2. SeaThru-NeRF

NeRF reconstructs the 3D space implicitly using a learnable continuous function. This function takes a point in 3D space and the angle of a ray passing through that point as input, and outputs the color and density of the point viewed from the perspective of the camera. This continuous function is approximated using a multilayer

perceptron. The architecture of NeRF assumes that 3D reconstruction takes place in a vacuum, so no light scattering or absorption occurs between the camera and the observed object. However, in reality, scattering media such as air and water cause the direction and strength of light rays to change as they travel from the object to the camera. SeaThru-NeRF addresses this by incorporating a model of image formation in scattering media [8], which adds medium color and density to the volume rendering process.

The color C at a point in the image plane is computed as follows by accumulating color densities along a ray $\mathbf{r}$ traveling in the 3D space from the camera:

$$C(\mathbf{r}) = \int_{t_n}^{t_f} T(t) \Big(\sigma^{\text{obj}}(t)\mathbf{c}^{\text{obj}}(t) + \sigma^{\text{med}}(t)\mathbf{c}^{\text{med}}(t) \Big)\, dt. \quad (1)$$

Here, $T(t)$ is the cumulative transmission probability that ray $\mathbf{r}$ travels along the parameterized variable t from the near bound of the domain t_n to the far bound t_f without colliding with other particles, $\sigma^{\text{obj}}(t)$ and $\mathbf{c}^{\text{obj}}(t)$ are the density and color of the object at t, and $\sigma^{\text{med}}(t)$ and $\mathbf{c}^{\text{med}}(t)$ are the density and color of the medium at t.

2.3. 3D Gaussian splatting

In 3D Gaussian splatting, objects in 3D space are explicitly reconstructed using 3D Gaussians, defined as

$$G(\mathbf{x}) = exp\left(-\frac{1}{2}(\mathbf{x})^T \Sigma^{-1}(\mathbf{x}) \right) \quad (2)$$

where Σ is a 3D covariance matrix. These Gaussians are generated on top of points comprising sparse point clouds derived from SfM pre-processing. Each 3D Gaussian is parameterized by its global position $\mathbf{p}_w$, quaternion $\mathbf{q}$, scale s, spherical harmonic coefficients h, and opacity α. During training, these Gaussians are projected onto the image plane of learning cameras, and the loss function minimizes the difference between rendered images (γ) and ground truth images (gt):

$$\arg \min_{\mathbf{p}_w, \mathbf{q}, s, h, \alpha} \mathcal{L} = (1-\lambda)\mathcal{L}_1 + \lambda \mathcal{L}_{D-SSIM}, \quad (3)$$

where

$$\mathcal{L}_1(gt, \gamma) = \frac{1}{N} \sum_{i=1}^{N} |gt_i - \gamma_i| \quad (4)$$

is the mean absolute error, and

$$\mathcal{L}_{D-SSIM}(gt, \gamma) = 1 - \frac{1}{N} \sum_{i=1}^{N} \text{SSIM}(gt_i, \gamma_i), \quad (5)$$

is a weighted D-SSIM loss with SSIM defined as in the SSIM evaluation metric, with the loss functions summed over the number of training images N.

The rendered pixel value γ is computed by accumulating contributions from all projected 3D Gaussians, weighted by their opacity α. The parameters $\{p_w, q, s, h, \alpha\}$ are updated via backpropagation to minimize $\mathcal{L}$. In the original paper, $\lambda = 0.2$ is used for balancing losses.

2.4. DUSt3R

DUSt3R is an end-to-end 3D reconstruction model that takes two RGB images as input and outputs two point maps $X^{1,1}, X^{2,1} \in \mathbb{R}^{W \times H \times 3}$ that map each pixel in the corresponding image to the 3D space and corresponding confidence maps $C^{1,1}, C^{2,1} \in \mathbb{R}^{W \times H}$. Here, $X^{n,m}$ represents the estimated 3D positions of pixels in the image captured by camera m and projected into the coordinate system of camera n, and $C^{n,m}$ encodes confidence values for each corresponding element in $X^{n,m}$.

The model architecture consists of shared ViT (Vision Transformer) encoders, a Transformer decoder with B blocks, and regression heads. Each image is processed independently by its ViT encoder to produce token features F_1 and F_2, which are initialized as G_0^1 and G_0^2. The Transformer decoder iteratively processes these tokens through B blocks. where each block applies self-attention, followed by cross-attention, and ends with a multilayer perceptron layer. Finally, the regression heads map the decoder outputs to the final predictions. This architecture enables end-to-end learning of 3D geometry and uncertainty estimation from image pairs.

2.5. Model Evaluation

To compare model performance, we use PSNR (Peak Signal-to-Noise Ratio), SSIM (Structural Similarity Index Measure), and LPIPS (Learned Perceptual Image Patch Similarity), which evaluate the quality of rendered images against ground-truth training images.

PSNR quantifies the pixel-wise fidelity between a reference image I and a rendered image I'. SSIM evaluates structural similarity between images by analyzing local luminance, contrast, and structure. It computes the average SSIM over a sliding window of size $H \times W$. LPIPS measures perceptual similarity by comparing feature maps from pre-trained CNNs (Convolutional Neural Networks) such as AlexNet and VGG across images.

While PSNR emphasizes pixel-level accuracy, SSIM captures structural and perceptual fidelity on a local level, and LPIPS reflects even higher-level human perception by using pre-trained CNNs to detect high-level perceptual differences. The combination of these metrics are widely used for evaluating 3D reconstruction and image generation tasks, as they balance quantitative precision with perceptual relevance.

3. RESULTS AND DISCUSSION

3.1. Dataset preparation

We used video footage recorded on March 29, 2023 by the underwater remote-operated vehicle ROV-A2 during an inspection of the containment vessel of Fukushima Daiichi Unit 1 reactor [9]. To prepare inputs for each model, still images were extracted from the original video. As shown in Fig. 2, the publicly available video integrates four camera feeds with hardcoded dates and times, making it incompatible with direct input to models. Thus, the top right camera feed was extracted, and regions containing text were manually cropped to gen-

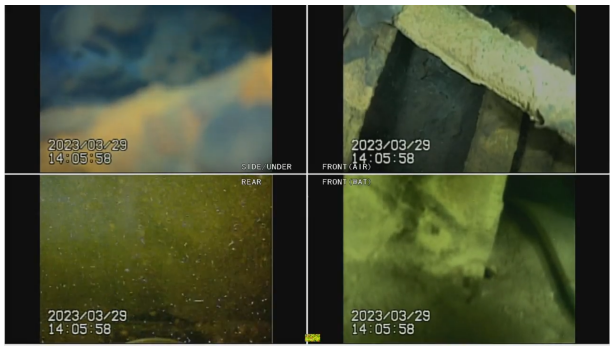

Fig. 2 Snapshot of image from video.

Fig. 3 Image trimmed from video.

Table 1 Video data.

Characteristic	Value
Resolution	700 × 400 px
Video length	2:24:31
Slice interval	0.3 s
Number of images used	27

erate clean inputs for all models, with a sample image shown in Fig. 3.

For SeaThru-NeRF and 3D-GS, only images for which COLMAP successfully estimated camera parameters and image positions prior to 3D reconstruction were used as training data. This led to variations in the number of input images across models due to preprocessing constraints. In contrast, DUSt3R simultaneously estimates camera parameters and image positions during 3D reconstruction, allowing all frames from the original video to be utilized for training. The specific recording conditions and cropping criteria used for input preparation are summarized in Table 1.

3.2. Reconstruction results

The 3D reconstruction results of SeaThru-NeRF, 3D-GS, and DUSt3R are visualized in Figs. 4, 5, and 6, respectively. According to Table 2, 3D-GS achieves superior performance across all evaluation metrics. SeaThru-NeRF produces continuous 3D reconstructions where structural features are recognizable only from viewpoints matching the training images. However, it fails to reconstruct structures at novel viewpoints due to limited generalization capabilities. 3D-GS generates discrete point clouds with high fidelity in texture and spatial distribution. Its superior performance in metrics is attributed to

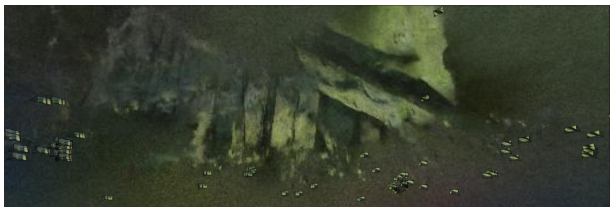

Fig. 4 SeaThru-NeRF reconstruction results.

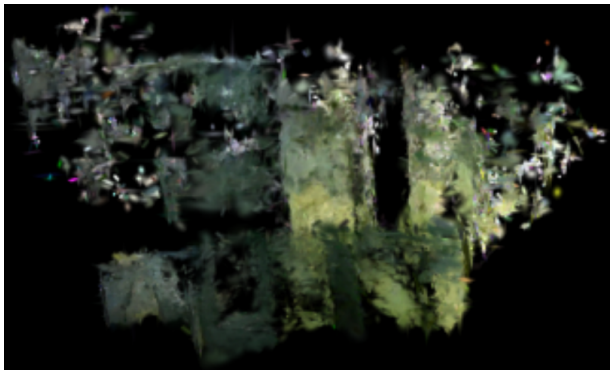

Fig. 5 3D-GS reconstruction results.

Fig. 6 DUSt3R reconstruction results.

Table 2 Evaluation results

	SeaThru-NeRF	3D-GS	DUSt3R
PSNR ↑	32.80	**36.95**	31.95
SSIM ↑	0.923	**0.966**	0.853
LPIPS ↓	0.15	**0.12**	0.24

its ability to capture complex underwater features such as sediment color gradients and object reflections, which are critical for accurate metric evaluation despite its discrete nature. DUSt3R achieves continuous and volumetric reconstructions with clear structural coherence. While it maintains spatial consistency across viewpoints, its performance in quantitative metrics is lower than 3D-GS due to the lack of explicit texture encoding in its reconstruction pipeline.

The dominance of 3D-GS in evaluation metrics stems from its ability to encode fine-grained details (e.g., sediment color gradients) directly into the 3D space via learned Gaussian splats, aligning closely with pixel values in training images. In addition, the point cloud-based approach of 3D-GS inherently preserves local texture information, which is critical for high PSNR/SSIM scores but less emphasized in volumetric methods like

DUSt3R or SeaThru-NeRF. This highlights a trade-off between perceptual realism, favoring continuous volumetric reconstructions, and metric accuracy, favoring discrete point cloud representations with explicit texture encoding. The methodology underscores the importance of task-specific design in 3D reconstruction pipelines, particularly for underwater environments where sedimentary features dominate visual fidelity.

The suboptimal qualitative 3D reconstruction results from SeaThru-NeRF and 3D-GS can be attributed to the following. Limited extraction of distinctive feature points occurred due to poor image quality, leading to degraded estimation of camera parameters and reduced training image counts. Also, particularly in the case of 3D-GS, Gaussians were likely placed at the positions of particulate matter/noise, resulting in disjunct reconstructions. This analysis underscores the critical role of dataset quality and preprocessing consistency in evaluating 3D reconstruction models for challenging underwater nuclear environments.

4. CONCLUSION

In this study, 3D reconstruction of the Fukushima Daiichi Unit 1 reactor containment vessel was conducted using video footage captured during inspections, employing SeaThru-NeRF, 3D-GS, and DUSt3R. Key findings include the following.

SeaThru-NeRF produces continuous reconstructions where structural features are discernible only from viewpoints aligned with training images, but fails to generalize to novel perspectives due to preprocessing limitations such as the reliance on COLMAP for camera parameter estimation. 3D-GS achieves discrete point cloud reconstructions with high fidelity in texture and spatial distribution. Its superiority in quantitative metrics stems from its ability to encode fine-grained details such as sediment color gradients directly into the 3D space, aligning closely with training image pixel values. However, faithful reproduction of suspended sediment led to reduced quality and discrete reconstructions. DUSt3R generates volumetric reconstructions that are continuous, spatially coherent, and structurally interpretable. While it performs well qualitatively, a lack of sediment reproduction and variance in lighting led to lower values during quantitative evaluation.

While 3D-GS outperformed other models in quantitative metrics, its discrete nature and preprocessing constraints highlight the need for improved data preparation workflows. Meanwhile, DUSt3R's strengths in spatial coherence suggest potential for hybrid approaches to achieve robust, task-specific reconstructions in underwater reactor environments. Future work will focus on benchmarking these models using standardized datasets and ground-truth point cloud data for reactor containment vessels.

ACKNOWLEDGEMENTS

This work was supported by the JAEA Nuclear Energy S&T and Human Resource Development Project (Grant Number: JPJA23O23813888). The authors would like to thank Wataru Suenaga for running experiments for the data included in this paper.

REFERENCES

[1] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis," *Communications of the ACM*, Vol. 65, No. 1, pp. 99–106, 2021.

[2] D. Levy, A. Peleg, N. Pearl, D. Rosenbaum, D. Akkaynak, S. Korman, and T. Treibitz, "SeaThru-NeRF: Neural Radiance Fields in Scattering Media," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 56–65, June 2023.

[3] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, "3D Gaussian Splatting for Real-Time Radiance Field Rendering," *ACM Transactions on Graphics*, Vol. 42, No. 4, pp. 1–19, July 2023.

[4] J. L. Schönberger and J.-M. Frahm, "Structure-from-Motion Revisited," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4104–4113, 2016.

[5] Y. Chen, S. Dong, X. Wang, L. Cai, Y. Zheng, and Y. Yang, "SG-NeRF: Neural Surface Reconstruction with Scene Graph Optimization," *European Conference on Computer Vision*, pp. 188–205, 2024.

[6] S. Wang, V. Leroy, Y. Cabon, B. Chidlovskii, and J. Revaud, "Dust3r: Geometric 3D Vision Made Easy," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20697–20709, 2024.

[7] M. Tancik, E. Weber, E. Ng, R. Li, B. Yi, T. Wang, A. Kristoffersen, J. Austin, K. Salahi, A. Ahuja, D. McAllister, J. Kerr, and A. Kanazawa, "Nerfstudio: A Modular Framework for Neural Radiance Field Development," *ACM SIGGRAPH 2023 Conference Proceedings*, pp. 1–12, August 2023.

[8] D. Akkaynak and T. Treibitz, "A Revised Underwater Image Formation Model," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[9] Tokyo Electric Power Company Holdings, Inc., *Photo and Video Library*, 2024.