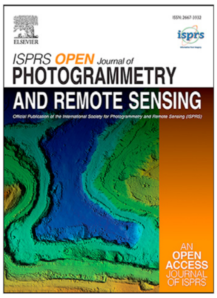
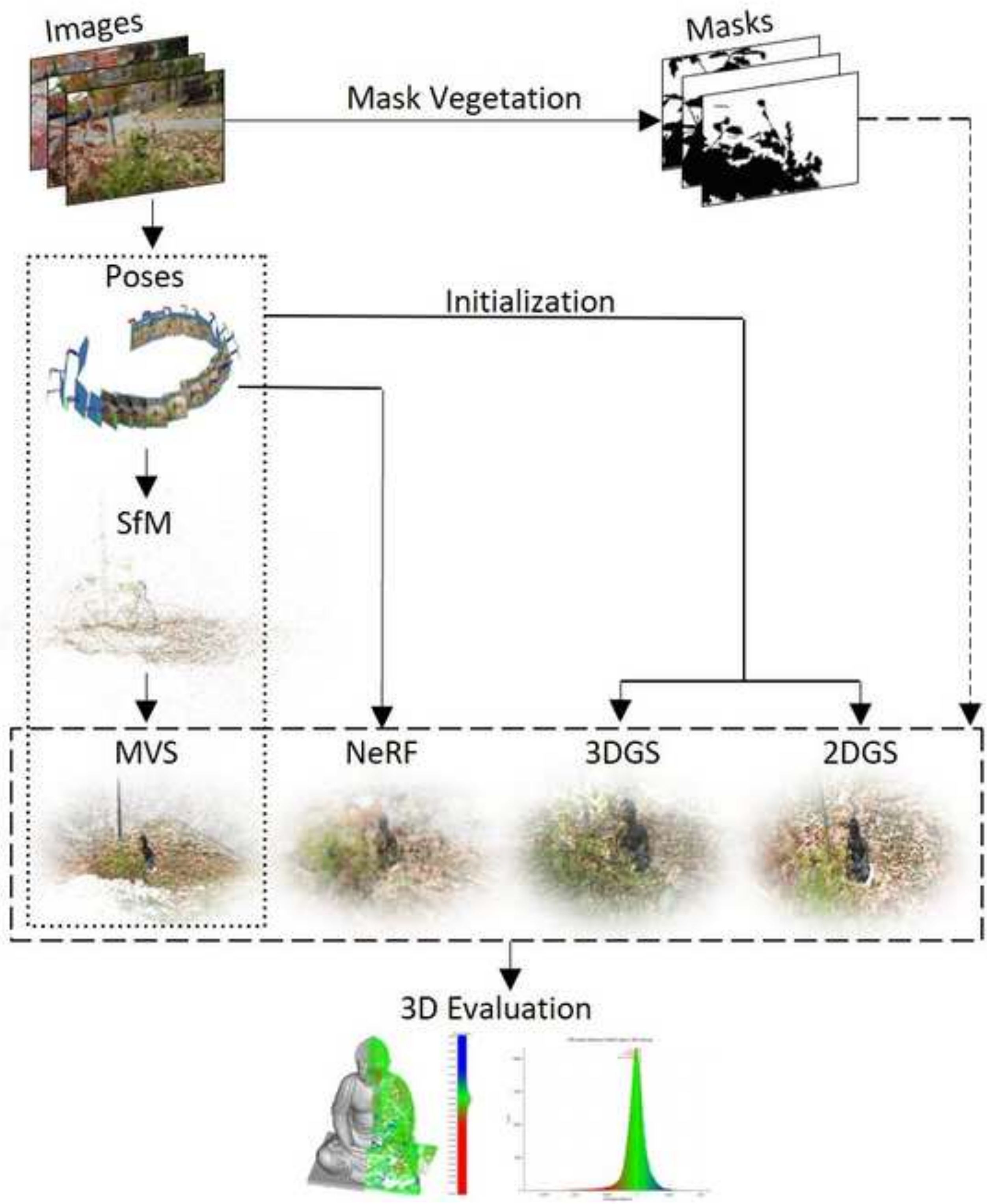


Seeing beyond vegetation: A comparative occlusion analysis between Multi-View Stereo, Neural Radiance Fields and Gaussian Splatting for 3D reconstruction

Ivana Petrovska, Boris Jutzi

PII:	S2667-3932(25)00008-0
DOI:	https://doi.org/10.1016/j.ophoto.2025.100089
Reference:	OPHOTO 100089
To appear in:	*ISPRS Open Journal of Photogrammetry and Remote Sensing*
Received date :	20 December 2024
Revised date :	7 March 2025
Accepted date :	31 March 2025

Graphical Abstract (for review)



Click here to view linked References

Highlights

Seeing Beyond Vegetation: A Comparative Occlusion Analysis between Multi-View Stereo, Neural Radiance Fields and Gaussian Splatting for 3D Reconstruction

Ivana Petrovska, Boris Jutzi

- Vegetation occlusions with various occlusion level for 3D reconstruction with traditional MVS and radiance field methods (NeRFs, 3DGS and 2DGS).
- MVS excels in pin-point accuracy, but fails under severe occlusions, while 2DGS achieves second best accuracy results outperforming NeRFs indicating a consistent performance across different occlusion scenarios.
- Comprehensive qualitative and quantitative 3D evaluation reporting accuracy and completeness.

Seeing Beyond Vegetation: A Comparative Occlusion Analysis between Multi-View Stereo, Neural Radiance Fields and Gaussian Splatting for 3D Reconstruction[⋆]

Ivana Petrovska[a,*], Boris Jutzi[a]

[a] Institute of Photogrammetry and Remote Sensing, Karlsruhe Institute of Technology (KIT), Karlsruhe, 76131, Germany

ARTICLE INFO

Keywords:
Multi-View Stereo
Neural Radiance Fields
Gaussian Splatting
Occlusion
Vegetation
3D Reconstruction
Point Clouds

ABSTRACT

Image-based 3D reconstruction offers realistic scene representation for applications that require accurate geometric information. Although the assumption that images are simultaneously captured, perfectly posed and noise-free simplifies the 3D reconstruction, this rarely holds in real-world settings. A real-world scene comprises multiple objects which obstruct each other and certain object parts are occluded, thus it can be challenging to generate a complete and accurate geometry. Being a part of our environment, we are particularly interested in vegetation that often obscures important structures, leading to incomplete reconstruction of the underlying features. In this contribution, we present a comparative analysis of the geometry behind vegetation occlusions reconstructed by traditional Multi-View Stereo (MVS) and radiance field methods, namely: Neural Radiance Fields (NeRFs), 3D Gaussian Splatting (3DGS) and 2D Gaussian Splatting (2DGS). Excluding certain image parts and investigating how different level of vegetation occlusion affect the geometric reconstruction, we consider Synthetic masks with different occlusion coverage of 10% (Very Sparse), 30% (Sparse), 50% (Medium), 70% (Dense) and 90% (Very Dense). To additionally demonstrate the impact of spatially consistent 3D occlusions, we use Natural masks (up to 35%) where the vegetation is stationary in the 3D scene, but relative to the view-point. Our investigations are based on real-world scenarios, one occlusion-free indoor scenario, on which we apply the Synthetic masks and one outdoor scenario, from which we derive the Natural masks. The qualitative and quantitative 3D evaluation is based on point cloud comparison against a ground truth mesh addressing accuracy and completeness. The conducted experiments and results demonstrate that although MVS shows lowest accuracy errors in both scenarios, the completeness manifests a sharp decline as the occlusion percentage increases, eventually failing under Very Dense masks. NeRFs manifest robustness in the reconstruction with highest completeness considering masks, although the accuracy proportionally decreases with increasing the occlusions. 2DGS achieves second best accuracy results outperforming NeRFs and 3DGS, indicating a consistent performance across different occlusion scenarios. Additionally, by using MVS for initialization, 3DGS and 2DGS completeness improves without significantly sacrificing the accuracy, due to the more densely reconstructed homogeneous areas. We demonstrate that radiance field methods can compete against traditional MVS, showing robust performance for a complete reconstruction under vegetation occlusions.

1. Introduction

3D reconstruction from multi-view images offers realistic representation of objects and scenes in applications that require accurate geometric representation. Capturing the spatial structure, point clouds are fundamental for depicting complex geometry in 3D computer vision and photogrammetry. Considering the camera poses along with the sparse point cloud estimated through Structure-from-Motion (SfM), Multi-View Stereo (MVS) involves triangulation of corresponding features in multiple images to calculate the 3D position of points in the scene. However, MVS relies on cross-view correspondence matching and triangulation which is time-consuming. Moreover, it restricts the completeness of the reconstructed point cloud, particularly for homogeneous areas, texture repetition and occlusions, assuming constant color to the viewing direction.

[⋆]

[*] Corresponding author.

✉ ivana.petrovska@partner.kit.edu (I. Petrovska); boris.jutzi@kit.edu (B. Jutzi)
(I. Petrovska); (B. Jutzi)
ORCID(s):

Addressing these limitations and representing detailed scene geometry, Neural Radiance Fields (NeRFs) (Mildenhall et al., 2021) reconstruct a scene through the weights of a coordinate-based neural network to predict the density and color at any given point from images and perspective transforms. As a position-dependent parameter, the density represents a probability of a point existing in scene space and belonging to an actual object, thus can be utilized to infer geometry at a queried position. Since the density field is a continuous implicit function, obtaining the explicit geometry of NeRFs as point cloud requires sampling rays from the training poses and capturing the 3D coordinates where a ray reaches its first major density peak for the rendering of depth maps. This restricts NeRFs geometric accuracy, as the depth is derived from the expected ray termination in the density field for which no constraints exist.

Revolutionizing radiance field reconstruction, 3D Gaussian Splatting (3DGS) (Kerbl et al., 2023) has emerged as an alternative to implicit scene representation, offering a competitive performance and providing an innovative solution for explicit point-based 3D geometry, where each point is represented as a Gaussian distribution. Employing images with corresponding poses and a sparse point cloud for initialization, the scene is reconstructed by optimizing Gaussian position, orientation, size, shape and appearance represented as spherical harmonics (SH). During optimization, the parameters of each individual Gaussian are updated via gradient descent over many iterations to best fit the training dataset. Compared to the time-consuming ray marching used in NeRFs, 3DGS optimizes on full images through a photometric loss in a single forward pass, omitting point computation in empty space. However, considering the mean of the Gaussians as scene geometry restricts the geometric accuracy, as the points lie behind the actual object surface. Inspired by this, 2D Gaussian Splatting (2DGS) (Huang et al., 2024) emerged as a promising alternative by representing scenes through 2D planar Gaussian disks which tightly align with object's surface, while simultaneously enhancing the geometry with depth-normal regularization.

Although the assumption that images are simultaneously captured, perfectly posed and noise-free simplifies the 3D reconstruction, this rarely holds in real-world settings where certain static or dynamic occlusions degrade the geometric representation. A real-world scene comprises multiple objects which obstruct each other and certain object parts are occluded, thus it can be challenging to generate a complete and accurate geometry. Being an inevitable part of our environment, we are particularly interested in vegetation that often obscures important structures, leading to incomplete reconstruction of the underlying features. The irregular and complex shape of vegetation and trees, especially dense foliage can introduce noise and distort the geometry of the reconstructed scene, reducing the overall accuracy and completeness.

In this contribution we analyze the geometry behind synthetic vegetation occlusions reconstructed by traditional MVS and radiance field methods, namely: NeRFs, 3DGS and 2DGS due to the different geometric representation. Selectively excluding certain parts of the images from influencing the reconstruction and investigating how different level of vegetation occlusions affect the geometric reconstruction, we consider Synthetic masks with different occlusion coverage of 10% (Very Sparse), 30% (Sparse), 50% (Medium), 70% (Dense) and 90% (Very Dense). To investigate how the initialization affects the overall 3D reconstruction in Gaussian Splatting (GS) methods, we use SfM and MVS point cloud without masks and MVS* with corresponding masks for each occlusion level respectively. Furthermore, to demonstrate the impact of spatially consistent 3D occlusions, we use Natural masks (up to 35%) where the vegetation is stationary in the 3D scene, but relative to the view-point. Our investigations are based on real-world scenarios; one occlusion-free indoor scenario, on which we apply the Synthetic masks and one outdoor scenario, from which we derive the Natural masks. The object of interest is placed behind vegetation to investigate how the methods allow to reconstruct the underlying geometry behind occlusions, thus investigating if radiance field methods can challenge traditional MVS for scenarios where the latter falls short. Consequently, the qualitative and quantitative evaluation is based on point cloud comparison against a ground truth mesh addressing accuracy and completeness metrics.

We provide the following contributions:

- We consider for the first time vegetation occlusions with different occlusion level for 3D reconstruction in 3DGS and 2DGS.
- We demonstrate that 2DGS with SfM initialization achieves second best accuracy results outperforming NeRFs indicating a consistent performance across different occlusion scenarios.
- We improve 3DGS and 2DGS completeness without significantly sacrificing the accuracy by leveraging MVS point cloud for initialization, due to the more densely reconstructed homogeneous areas.
- We show that radiance field methods: NeRFs, 3DGS and 2DGS can compete against traditional MVS showing robust performance for a complete reconstruction under severe vegetation occlusions.

- We provide a comprehensive qualitative and quantitative 3D comparison among MVS, NeRFs, 3DGS and 2DGS reporting accuracy and completeness.

The contribution is organized as follows. In Section 2 an overview of current NeRF and GS methods is presented, in Section 3 the processing of SfM and occlusion mask generation along with the used 3D reconstruction methods and evaluation metrics is given, information about the dataset and ground truth along with implementation details is provided in Section 4, in Section 5 the qualitative and quantitative results for accuracy and completeness are presented, the discussion is laid out in Section 6 and Section 7 reports the concluding insights.

2. Related Work

In the following, we briefly discuss related work in the area of 3D reconstruction with radiance fields, tackling 3D scene representation in occlusion settings. For this purpose, we first summarize approaches regarding NeRFs geometric reconstruction in Section 2.1, followed by recent research and development in GS methods in Section 2.2.

2.1. Neural Radiance Fields

With their foundation established with Scene Representation Networks (SRN) (Sitzmann et al., 2019), NeRFs (Mildenhall et al., 2021) estimate position-dependent density and view-dependent color for each point by training a fully connected neural network on images and associated camera poses. However, NeRFs assume controlled conditions and bounded scenes and can be computationally demanding as rendering each ray requires querying the network hundreds of times. Inspired by these drawbacks, Mip-NeRF (Barron et al., 2021) resolves the anti-aliasing issue by casting conical frustums instead of rays allowing scene reconstruction on different scales, while Mip-NeRF360 (Barron et al., 2022) tackles the challenges presented by unbounded real-world scenes with unconstrained camera orientation. Advancing NeRFs once again, Instant Neural Graphic Primitives (Instant-NGP) (Müller et al., 2022) overcome the issue of computational efficiency using a small neural network augmented by multi-resolution hash encoding and binary occupancy grid, skipping empty space within the scene.

Leveraging a point cloud to generate depth complementing the image information, LiDeNeRF (Wei et al., 2024) enhances scene geometry and appearance. Point-NeRF (Xu et al., 2022) and Points2NeRF (Zimny et al., 2022) use neural point clouds to model a point-based radiance field, but only in one scale. Effectively overcoming this, PointNeRF++ (Sun et al., 2023) unifies point clouds and NeRFs to adapt to variable point densities and empty regions. Neuralangelo (Li et al., 2023) combines multi-resolution 3D hash grids and neural surface rendering to achieve superior results in recovering dense geometry from multi-view images, enabling highly detailed scene reconstruction. Plenoxels (Fridovich-Keil et al., 2022) and DVGO (Sun et al., 2022) directly replace the neural network with a dense voxel grid, performing volume rendering on the interpolated 3D features. However, as the scene size increases, using voxel grid representation leads to an exponential increase in memory. NeRFBK (Karami et al., 2023) initiated an advance in evaluating the 3D reconstruction by NeRFs for objects with different surface properties, captured indoors and outdoors.

Learning a radiance field under occlusions, NeRF-W (Martin-Brualla et al., 2021), Ha-NeRF (Chen et al., 2022) and SF-NeRF (Lee et al., 2023) extend NeRFs capabilities in uncontrolled real-world environments by decomposing the scene into transient and static components. However, they heavily rely on highly accurate camera extrinsic parameters. RobustNeRF (Sabour et al., 2023) offers robust optimization with distractors that are not persistent throughout the capture session, but yields poorer reconstruction on clean scenes. By the assumption that the occlusions are in front of the object, Occlusion-Free NeRF (Zhu et al., 2023) predicts the occluding likelihood based on the weights along the ray. If the expected ray termination depth from the camera and that from the other end of the ray show significant difference, the ray likely passes through some foreground occlusion. Thus, by aggregating information from different view-points the occlusions are eliminated. Nevertheless, the method is primarily designed for novel view synthesis, as the occlusions comprise a small image portion and the datasets have a limited number of images, making it unsuitable for 3D reconstruction tasks.

2.2. Gaussian Splatting

In contrast to the continuous volumetric field of NeRFs, 3DGS (Kerbl et al., 2023) leverages a point-based representation associated with 3D Gaussian attributes initialized on a sparse point cloud from the camera poses, imposing pre-processing and calculation steps. COLMAP-Free 3DGS (Fu et al., 2023) eliminates the need of camera poses by processing the images with known intrinsics in a sequential manner to progressively grow the Gaussians. Although random initialization achieves accuracy similar to SfM (Dai et al., 2024; Foroutan et al., 2024) in synthetic scenarios, for real-world scenes 3DGS heavily relies on a good point cloud for initializing the Gaussians. Improving the robustness to initialization, 3DGS-MCMC (Kheradmand et al., 2024) aligns the 3D Gaussians with the principles of Markov Chain Monte Carlo sampling. The densification is reformulated to employ a relocalization scheme for moving low-opacity Gaussians to the locations of Gaussians with high opacity, thus managing their number effectively and reducing computational time. MVPGS (Xu et al., 2024b) initializes the Gaussians on a learning-based MVS with monocular depth priors as MVS performs poorly in homogeneous areas. Additionally, to facilitate geometric convergence, a view-consistent depth loss between the 3D Gaussians and the MVS dense point cloud is used. However, the evaluation focuses on novel views, lacking 3D geometric accuracy and completeness.

With a view to more accurately represent surface geometry, SuGaR (Guédon and Lepetit, 2024), NeuSG (Chen et al., 2023) and GeoGaussian (Li et al., 2024b) propose a regularization term that flattens the Gaussians by minimizing the smallest scale to encourage thin Gaussians to align with intricate surfaces, which comes at a computational cost. Similarly, 2DGS (Huang et al., 2024) collapses the 3D ellipsoids into a set of 2D Gaussian surfels, enforced with depth-normal consistency regularization. 3D-HGS (Li et al., 2024a) proposes to split each Gaussian into two halves and assign different opacity to each of them. Although the method demonstrates remarkable results in novel scene generation, it still encounters challenges with geometry reconstruction in homogeneous areas. GS2Mesh (Wolf et al., 2024) instead of applying geometric optimization directly on the Gaussians using their position uses a pre-trained stereo model as a geometric prior to extract the depth, enhancing surface representation.

All of the above mentioned methods assume that images are simultaneously captured with unobstructed line of sight which is very unlikely in real-world setup. SpotlessSplats (Sabour et al., 2024) exploit semantic features to effectively identify transient distractors without any explicit supervision. Nonetheless, when distractors and non-distractors of the same semantic class are present and in close proximity, they may not be distinguished and thin structures can be missed. WildGaussians (Kulhanek et al., 2024) and Splatfacto-W (Xu et al., 2024a) extend GS to uncontrolled in-the-wild settings containing occluders, under-performing in challenging scenarios where occlusions are present in nearly all training images.

In summary, NeRFs and GS subsequent methods are primarily focused on the task of novel view synthesis and take into account small occlusions. Our investigations differ from state-of-the-art because we consider vegetation and analyze the geometry behind severe occlusions, comprising up to 90% of the image. Moreover, we initialize the Gaussians on a MVS point cloud without the advantage of priors and further geometric improvements for a fair analysis on the impact of the initialization to the 3D geometric reconstruction in vegetation occlusion scenarios.

3. Methodology

As illustrated in Figure 1, the principles of SfM for camera pose estimation and sparse point cloud reconstruction are introduced in Section 3.1. Subsequently, in Section 3.2 the 3D reconstruction methods are summarized, the occlusion mask generation is laid out in Section 3.3, while the evaluation metrics are presented in Section 3.4.

3.1. Structure-from-Motion (SfM)

The intrinsic and extrinsic camera parameters along with a sparse point cloud are estimated through SfM (Schonberger and Frahm, 2016) which relies on extraction of distinctive features within an image sequence from a set of overlapping images captured from different position. The extracted image features are then matched against each other to establish correspondences between the individual images to create a scene graph enforcing geometric consistency. Refining the estimates of extrinsic camera parameters and 3D points, a global bundle adjustment is employed. The images along with estimated camera poses are necessary for training NeRFs, while 3DGS and 2DGS additionally require the sparse SfM point cloud for initialization.

3.2. 3D Reconstruction Methods

In the following we describe the applied 3D reconstruction methods. The MVS dense scene representation is briefly summarized in Section 3.2.1, in Section 3.2.2 the geometric reconstruction and point cloud extraction of NeRFs is described, while the principles of 3DGS and 2DGS are explained in Section 3.2.3 and Section 3.2.4 accordingly.

3.2.1. Multi-View Stereo (MVS)

Based on the camera poses and sparse point cloud from SfM, MVS (Schönberger et al., 2016) involves the triangulation of corresponding features in multiple images to calculate the 3D position of points in the scene through

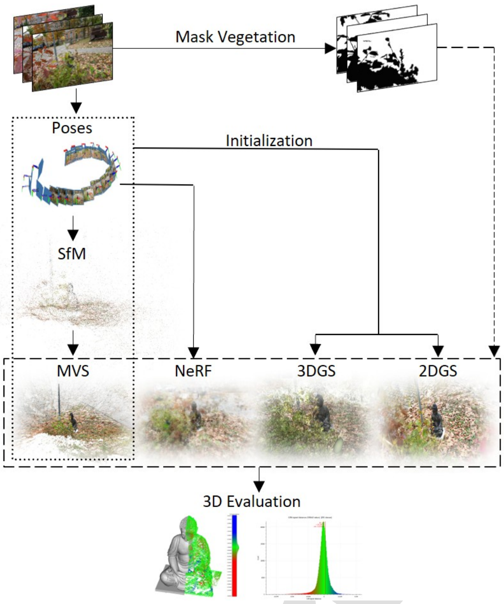


Figure 1: Overview of the processing steps. The camera poses and sparse point cloud are estimated through SfM (Section 3.1), followed by dense MVS reconstruction (Section 3.2.1). Additionally, we consider occlusion masks (Section 3.3) by excluding the vegetation from the 3D reconstruction. The images along with the poses are input for NeRF (Section 3.2.2). The SfM and MVS point cloud without and MVS with corresponding masks are used for initialization in 3DGS (Section 3.2.3) and 2DGS (Section 3.2.4). We evaluate the reconstructed point clouds without and with masks addressing accuracy and completeness (Section 3.4).

pixel-wise computation of depth information in an image. In a final step, the geometric consistent depth maps are fused into a dense point cloud. Additionally, we consider the occlusion masks in MVS dense reconstruction by setting a value of 0 for the masked parts in the input images. The MVS point cloud without and with Synthetic masks are used for initialization in 3DGS and 2DGS.

3.2.2. Neural Radiance Fields (NeRFs)

NeRFs (Mildenhall et al., 2021) implicitly represent a 3D scene by a trainable continuous function. The multi-layer perceptron (MLP) neural network encodes the density σ at 3D point $x = (x, y, z)$ and the color $c = (r, g, b)$ emitted from this point in viewing direction $d = (\theta, \phi)$. The scene is reconstructed by querying radiance along rays computed as a volume integral estimated by drawing samples along the ray, evaluating their density and color, then accumulating values. The neural network is trained by minimizing the image reconstruction loss over training views through gradient descent. We also apply the occlusion masks which indicate areas that should have zero density at coordinates corresponding to masked image pixels.

In the original Instant-NGP (Müller et al., 2022) implementation we extract the point cloud by voxelizing the density field. Nevertheless, filtering with global density thresholds yields noisy and incomplete reconstruction and depends on the input data (Oechsle et al., 2021; Petrovska et al., 2023; Jäger et al., 2023), thus we apply a 3D density-gradient based Canny edge detection filter. The gradient calculation based on the position-dependent density begins with smoothing the density field to suppress noise. This is followed by gradient magnitude relative thresholds, facilitating the identification of edges. Finally, a hysteresis method tracks strong edges while simultaneously suppressing weaker ones. The density gradients are determined independently of the absolute magnitude of density values, enabling extraction of edges in regions with lower density within the field (Jäger and Jutzi, 2023). The total Canny gradient $\Delta_{\delta,\text{Canny}}$ is given by:

$$\Delta_{\delta,\text{Canny}} = \sqrt{G_{\delta,x}^2 + G_{\delta,y}^2 + G_{\delta,z}^2} \tag{1}$$

where $G_{\delta,x}$, $G_{\delta,y}$ and $G_{\delta,z}$ are the density gradients in direction x, y and z in the density field.

NeRFs predict the continuous density field in the entire 3D space, consequently inside the object. We remove these points by approximating the visibility of the exported point cloud from a given view-point. We set up a virtual camera relative to the centroid of the point cloud. Then, points that are not visible from the specified camera view-point are filtered. We render five different camera views with regard to a given radius, that determines the spherical region around a given view-point within which points are considered for visibility. We then merge all five point cloud parts.

3.2.3. 3D Gaussian Splatting (3DGS)

The scene is represented as a set of 3D Gaussians (Kerbl et al., 2023), where each Gaussian has its mean μ and anisotropic covariance matrix Σ, parameterized by a scaling vector $s \in \mathbb{R}^3$ and a quaternion $q \in \mathbb{R}^4$ encoding the rotation. Each Gaussian is associated with opacity $\alpha \in [0,1]$ and spherical harmonics (SH) describing the color in the radiance field. Given a set of posed images, the Gaussian primitives are initialized on a point cloud. During optimization, the Gaussian parameters are optimized by cloning, splitting and culling to match the input images, minimizing the photometric loss which is a combination of *D-SSIM* term and $\mathcal{L}_1$ loss for per-pixel color differences computed between rendered $\hat{C}$ and ground truth C images with $\lambda_{\text{D-SSIM}} = 0.2$:

$$\mathcal{L}_{\text{3DGS}} = (1 - \lambda_{\text{D-SSIM}})\|\hat{C} - C\|_1 + \lambda_{\text{D-SSIM}}\,\text{D-SSIM}(\hat{C}, C) \tag{2}$$

We consider the occlusion masks (Section 3.3) in the loss function restricting the gradients to learn only from the unmasked pixels:

$$\mathcal{L}_{\text{3DGS\_MASK}} = (1 - \lambda_{\text{D-SSIM}})\|\hat{C} - C_M\|_1 + \lambda_{\text{D-SSIM}}\,\text{D-SSIM}(\hat{C}, C_M) \tag{3}$$

Due to the method's explicit representation, the Gaussian mean is considered object geometry as point cloud. For visualization purposes, we color code the point cloud by converting the SH back to RGB values.

3.2.4. 2D Gaussian Splatting (2DGS)

A 2D splat (Huang et al., 2024) is characterized by its central point p_k, two principal tangential vectors t_u and t_v, rotation matrix $R = [t_u, t_v, t_w]$ and a scaling vector $S = (s_u, s_v)$ controlling the size and shape of the 2D Gaussian disks. Each 2D Gaussian primitive has opacity α and view-dependent appearance c parameterized by SH. The parameters are learned through gradient descent between rendered and ground truth images. However, as optimizing solely with photometric loss can lead to noisy reconstruction, two additional regularization terms are introduced: depth distortion $\mathcal{L}_D$ to concentrate the 2D primitives along the rays by minimizing the distance between the ray-splat intersections and normal consistency $\mathcal{L}_\mathcal{N}$ which minimizes discrepancies between the rendered normal map and the gradient of the rendered depth:

$$\mathcal{L}_{\text{2DGS}} = \mathcal{L}_{\text{3DGS}} + \alpha\mathcal{L}_D + \beta\mathcal{L}_\mathcal{N} \tag{4}$$

Following the original implementation, $\alpha = 1000$ for bounded scenes, $\alpha = 100$ for unbounded scenes and $\beta = 0.05$ for all scenes. Subsequently, considering the occlusion masks (Section 3.3) the total loss is minimized guiding the algorithm not to propagate for the masked image pixels:

$$\mathcal{L}_{\text{2DGS\_MASK}} = \mathcal{L}_{\text{3DGS\_MASK}} + \alpha\mathcal{L}_D + \beta\mathcal{L}_\mathcal{N} \tag{5}$$

We consider the center of the Gaussian disks as point cloud representation and color code the point cloud by converting the SH back to RGB values.

Table 1
Synthetic occlusion masks overview. Starting from grass to dense forest, we consider five Synthetic mask variations tackling Very Sparse, Sparse, Medium, Dense and Very Dense foliage with masked pixels of 10, 30 50, 70 and 90% respectively.

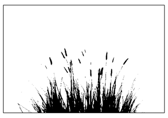
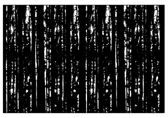
Type of mask	Synthetic				
	Very Sparse	Sparse	Medium	Dense	Very Dense
RGB Mask Preview					
Binary Mask Preview					
Description	Cluster of tall grass	Thin branches with leaves	Trees with sparse foliage	Forest with slender trees	Forest with thick trunks
Percentage	10%	30%	50%	70%	90%

Table 2
Natural occlusion masks overview. The Natural masks depict vegetation occlusions which exist and are static in the scene, but change their position in the images according to the view-point. The vegetation acts an occlusion when in front of the object with average of 35% masked pixels.

Type of mask	Natural
RGB Mask Preview	
Binary Mask Preview	
Description	Bush with thin leaves
Percentage	35%

3.3. Occlusion Masks

To investigate how different level of vegetation occlusions affect the geometric reconstruction, we consider two types of binary masks $M(u, v) \in [0, 1]$ (Table 1 and 2). All masks will be publicly available.

Synthetic. We first generate RGB images from text prompt with Stable Diffusion in a Hugging Face environment[1] which we place on an empty image with the same resolution as our input images. We convert it into a binary mask by filtering through intensity values, namely: if gray value is >= 170 set mask to 1, otherwise set to 0. We consider five synthetic mask variations tackling Very Sparse, Sparse, Medium, Dense and Very Dense foliage starting from grass to forest environments. The Very Sparse and Sparse masks have zero-valued pixel coverage of 10.37% and 30.37% accordingly, Medium masks 50%, Dense 70.84% and the Very Dense 88.94%. For simplicity and readability we round the percentages to 10, 30, 70 and 90 respectively. Starting from sparse tall grass for the Very Sparse masks, the Sparse masks describe thin tree branches with small leaves. The Medium masks consist of tall trees with sparse foliage. The Dense masks depict forest with tall slender trees covered with vegetation, while for the Very Dense masks, thicker trunks and dense forest with a mix of green foliage can be identified (Table 1). We apply the same synthetic mask to all images of an occlusion-free indoor scenario to have control over the mask percentage and ensure it remains constant on all images.

Natural. For this purpose we use an outdoor scenario where the vegetation occlusions exist and are static in the scene, but change their position in the images according to the view-point. We annotate the vegetation using the

[1] https://huggingface.co/spaces/stabilityai/stable-diffusion

Segment Anything Model (SAM) (Kirillov et al., 2023) in Roboflow[2], then convert the segmentation polygons into binary masks for each image. The vegetation in form of a bush with thin leaves attached on branches has an average of 33.93% of foreground occlusion. For consistency, we round this percentage to 35% (Table 2).

3.4. 3D Evaluation

Image-based metrics only evaluate the prediction quality via a 2D projection, which leads to loss of accuracy information in volumetric space. This makes them misleading indicators of performance for many real-world applications since the data is captured in unstructured environments, justifying a 3D evaluation. The evaluation on the level of point clouds requires a co-registration with a ground truth, first with coarse alignment picking equivalent point pairs in both entities, followed by fine registration using the Iterative Closest Point (ICP) (Besl and McKay, 1992), which finds an optimal rigid transformation to align two point sets by minimizing the distance between the points. To evaluate the geometry of the reconstructed point clouds, we report accuracy and completeness.

Accuracy. Accuracy quantifies how dispersed the reconstructed point cloud is in relation to the ground truth. We adopt cloud-to-mesh distances, which compute the displacements between each point in the compared point cloud and the nearest facet in the ground truth mesh through Euclidean distance. The orthogonal (signed) distance from the point to the nearest triangle plane or nearest edge is taken. We report Mean Error (Mean), Standard Deviation (SD) which measures the spread of the points around the Mean and Root Mean Square Error (RMSE) used to measure distances between predicted and actual values:

$$Mean = \frac{\sum_{i=1}^{n}(d_i)}{n} \tag{6}$$

$$SD = \sqrt{\frac{\sum_{i=1}^{n}(d_i - \overline{d})^2}{n-1}} \tag{7}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(d_i)^2}{n}} \tag{8}$$

where d_i is the distance between each point in the compared point cloud and the nearest facet of the ground truth mesh, $\overline{d}$ denotes the mean value of the distances and n is the total number of points in the compared point cloud.

Completeness. Completeness (*Cpl*) measures to what extent the ground truth surface is covered and is calculated as the ratio of the number of covered points to the total number of points in the ground truth. All points in the ground truth within a specified threshold distance of an estimated point contribute to the completeness. For this purpose we convert the mesh into a point cloud by subsampling points on the mesh. A higher score indicates higher completeness, conditioned by the number of points (*Npts*) and is reported as a percentage.

$$Cpl = \frac{\sum_{i=1}^{N}}{N} 1(d_i \leq t) \tag{9}$$

where the total number of points in the ground truth point cloud $N = 10M$, d_i denotes the distance between a ground truth point and the closest point in the compared point cloud, the predefined distance threshold $t = 5mm$ and $1(d_i \leq t)$ is an indicator that equals 1 if $d_i \leq t$ (the point is covered) and 0 otherwise.

4. Experiments

After a brief description of the dataset in Section 4.1 and ground truth mesh generation in Section 4.2, the implementation details for each method are outlined in Section 4.3.

[2] `https://roboflow.com/`

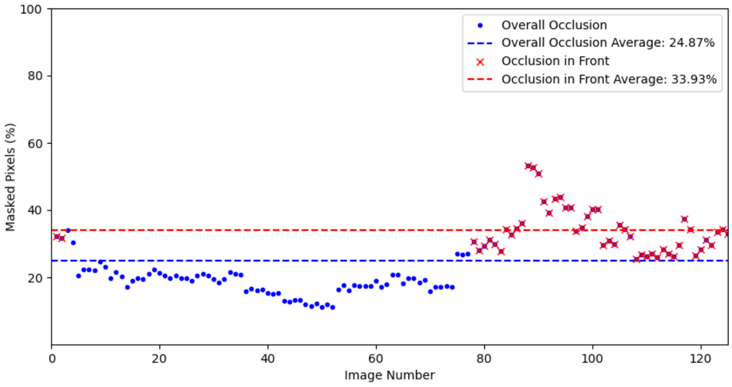


Figure 2: Natural masks where the vegetation occlusions are present and static in the 3D scene, but their position in the images is relative to the viewing direction. Out of 125 images (blue dots), vegetation acts as an foreground occlusion in 50 images (red x marks) as it's in front of the object.

4.1. Dataset

Our investigations are based on two real-world scenarios (Figure 3) from the STELLA[3] dataset, namely a bounded indoor occlusion-free (*Original*) and unbounded outdoor (*Vegetation*). The object whose geometry ought to be evaluated is 0.7m tall Buddha statue (further on referred to as object). The images are captured using a Nikon D810 SLR digital camera with a 36MP camera sensor, image resolution of 7360×4912px, 20mm focal length and f/8 aperture size. Each scenario consists of 125 images captured on a circular trajectory for full object coverage (Petrovska and Jutzi, 2024). Due to memory limitation and efficiency the images are downsampled to 1840x1228px and converted into .png format for lossless compression, same as the occlusion masks. The Synthetic masks are applied to all images in *Original*, while the vegetation is present in *Vegetation* from which the Natural masks are derived. However, in 50 images the vegetation acts as a foreground occlusion being in front of the object with average of 33.93% masked pixels (Figure 2).

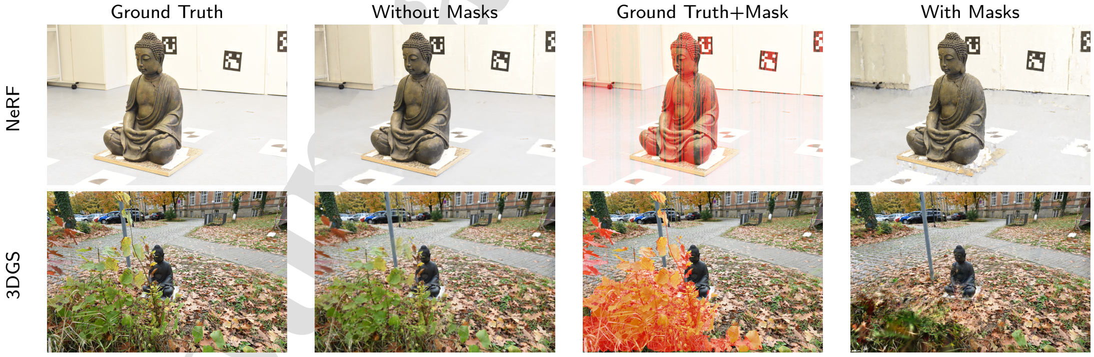


Figure 3: Novel views of NeRF for *Original* (above) and 3DGS for *Vegetation* (below). The highlighted parts are the masked image pixels. NeRF shows reliable reconstruction without masks and with Dense masks (70%) where the rendering quality is lower due to less image information. The 3D Gaussian splats which act like blobs in space with view-dependent appearance are visualized without masks and with Natural masks, leading to occlusion-free reconstruction.

[3] https://github.com/sqirrel3/STELLA

Table 3
For GS initialization we use SfM and MVS without masks and MVS* with corresponding Synthetic masks for each occlusion level accordingly. 3DGS and 2DGS are initialized on the point cloud of the whole scene (Scene points). As we evaluate the geometry behind occlusions, we remove redundant points to keep just the object (Object points) and report the scores for MVS in Table 4.

Scenario (Original)	Initialization	Scene points	Object points
Without Masks	SfM	42.580	31.866
	MVS	3.219.397	845.456
Very Sparse Masks	MVS*	2.870.659	739.936
Sparse Masks	MVS*	1.528.457	554.636
Medium Masks	MVS*	93.442	73.937
Dense Masks	MVS*	10.541	5.407
Very Dense Masks	MVS*	42	0

4.2. Ground Truth

As ground truth, we use Structured Light Imaging (SLI) mesh with 0.1mm accuracy captured with stereoSCAN scanning device whose projector is placed between two digital cameras at a fixed distance from each other. The light source projects a coded pattern of parallel light stripes onto the object and the cameras capture these patterns from a known position, resulting in a specific sequence of gray values for each pixel of an image, from which the range can be calculated. The coordinates of the object points are then triangulated from the intrinsic and extrinsic camera parameters and the image coordinates. During acquisition, the object is placed on a turntable with rotations automatically controlled by a workstation (Petrovska and Jutzi, 2024).

4.3. Implementation Details

MVS. We use COLMAP[4] as it implements a full end-to-end SfM and MVS pipeline. For the incremental SfM the exhaustive matcher is used during pose estimation through SIFT (Lowe, 2004) for feature extraction and matching. The images with corresponding poses are input to NeRF, while for 3DGS and 2DGS the sparse SfM point cloud and MVS without masks are used for initialization. Additionally, we use the MVS* with corresponding Synthetic masks for each occlusion level respectively (Table 3).

NeRF. As NeRF representative we employ Instant-NGP with hashmap size of 2^{19} with a resolution level of 16, each level containing 2 dimension of features, consistent with the original implementation. The Canny filter is applied with a standard derivation of 0.1 and relative thresholds between 0.01 and 0.001 depending on the scene.

3DGS & 2DGS. We train with the default hyperparameters with learning rates of 0.0025 for SH, 0.005 for scaling operations, 0.001 for rotation transformations and 0.0002 threshold of positional gradient for densification. The opacity values are set close to zero every 3.000 iterations to remove splats with opacity lower than 0.05.

All experiments regarding training and evaluation are performed on a Nvidia RTX3090 GPU and Intel i9 CPU with 32GB RAM. NeRFs and GS are trained for 30.000 iterations with every 8th frame taken for test during training. The accuracy evaluation is performed in CloudCompare[5], while the completeness calculation in our python script.

5. Results

To evaluate how the occlusion masks affect the geometric reconstruction, we compare the point clouds reconstructed by MVS, NeRF, 3DGS and 2DGS and argue which method represents the geometry behind occlusions more accurately and reliably. As initialization for 3DGS and 2DGS, we use the SfM without masks and MVS without and with Synthetic masks (MVS*) to investigate the influence of the initial point cloud to the geometric representation. Note that MVS* represents the whole scene and for evaluation we keep just the object (Table 3). We report qualitative,

[4] https://github.com/colmap/colmap
[5] https://github.com/CloudCompare/CloudCompare

where the 3D reconstructions as point clouds are visualized (Figure 4, 5, 8 and 9) as well as the cloud-to-mesh errors (Figure 6, 7, 10 and 11) and quantitative results to numerically analyze the accuracy and completeness for each method separately (Table 4 and 5).

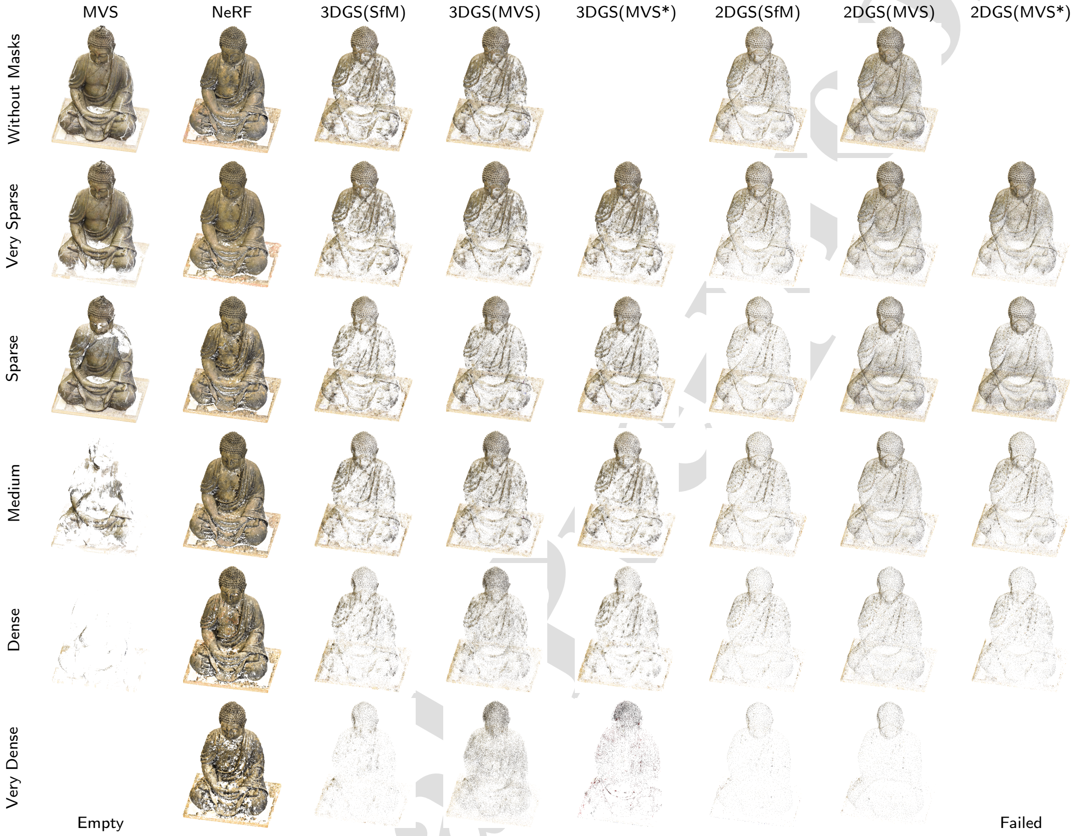


Figure 4: Point cloud reconstruction for *Original* using Synthetic masks with different occlusion coverage. MVS completeness is significantly affected. With Very Dense masks, the scene has only 42 background non-object points used for GS initialization. NeRF exhibits highest point coverage indicating robustness in the reconstruction. For GS initialization, we use SfM and MVS without and MVS* with corresponding masks for each occlusion level respectively. The completeness proportionally decreases with masks however, MVS and MVS* initialization improves the object surface representation. Initialized on MVS* with Very Dense masks, 2DGS reconstruction fails.

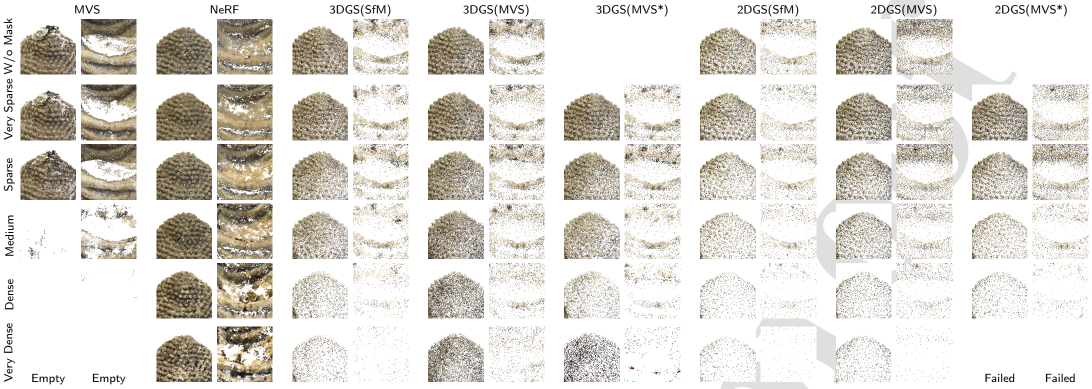


Figure 5: Enlargements of complex geometry (head spikes) and homogeneous areas (lap) for *Original*. MVS manifests black and white artifacts and gaps in the reconstruction. With Very Dense masks, the scene has only 42 background non-object points used for GS initialization. NeRF achieves higher object completeness and better reconstructs the underlying geometry behind occlusions. In 3DGS and 2DGS the complex geometry is already reliably reconstructed with SfM initialization, however MVS without and MVS* with masks for each occlusion level respectively populate the homogeneous areas, increasing the completeness. Initialized on MVS* with Very Dense masks, 2DGS reconstruction fails.

5.1. Multi-View Stereo

Although MVS reliably reconstructs the object and achieves a sharp result for *Original* without masks (Figure 4), we can notice black and white artifacts, especially on the spikes on the head and left shoulder as well as noise in the lap that is sparsely reconstructed (Figure 5). These points have high error values (Figure 6 and 7). Nevertheless, it shows lowest RMSE of 1.43mm and highest completeness of 97.35% (Table 4). Although MVS exhibits the best accuracy among the methods, the point coverage proportionally drops with occlusions until Medium masks when the decline is sharp (Figure 13), eventually failing to reconstruct the object under Very Dense masks. With Very Sparse masks it maintains the same accuracy of 1.43mm, however the completeness is affected and declines to 83.62% due to the gaps in the lower object part and the plate corresponding to the vegetation occlusions. Increasing the occlusions to 30% leads to slightly lower RMSE on account of the number of points which drops to 554.636. Considering Medium masks, MVS error increases to 2.72mm and the object is partially reconstructed resulting in 3x lower completeness of 30.49%. Further increasing the mask percentage to 70% leads to better accuracy of 1.67mm on account of point coverage as the object has only 5.407 points and only some edges are reconstructed. When the mask coverage is severe 90%, MVS fails to reconstruct the object. The point cloud contains only 42 scattered background points not belonging to the object, which are used for initialization in 3DGS and 2DGS. In *Vegetation*, despite failing to reconstruct the occluded object parts (Figure 8), it still exhibits the best results with highest correspondence with the ground truth (Figure 10 and 11) with RMSE of 3.23mm and 2.77mm without and with masks respectively (Table 5). This is complemented by high completeness of above 75%, implying a well-balanced trade-off between precision and surface coverage outperforming radiance field methods. Nevertheless, it fails to capture the intricate geometric details on the head (Figure 9).

5.2. Neural Radiance Fields

NeRF point clouds in both scenarios (Figure 4 and 8) capture the overall spatial arrangement of the object completely while being able to capture complex geometric details like the spikes on the head (Figure 5 and 9), even behind occlusions. The lap is also reconstructed with more points compared to the other methods. For *Original* without masks, it exhibits second best accuracy with RMSE of 3.71mm and high completeness of just below 95% (Table 4). The performance proportionally drops with increasing the occlusions, but not as steep as MVS. NeRF achieves best completeness under Very Sparse masks with 94.35% and second best accuracy with RMSE of 3.26mm. Noise and artifacts are noticeable in the lap (Figure 7) as that part is occluded by vegetation corresponding to the Synthetic mask. Maintaining moderate accuracy, it reaches high surface coverage under Sparse occlusions with 94.33%. With Medium occlusions, the accuracy decreases slightly to 4.12mm, but the completeness is the highest among the methods and almost the same as with Very Sparse masks. Under Dense masks the accuracy drops to 7.64mm, but despite

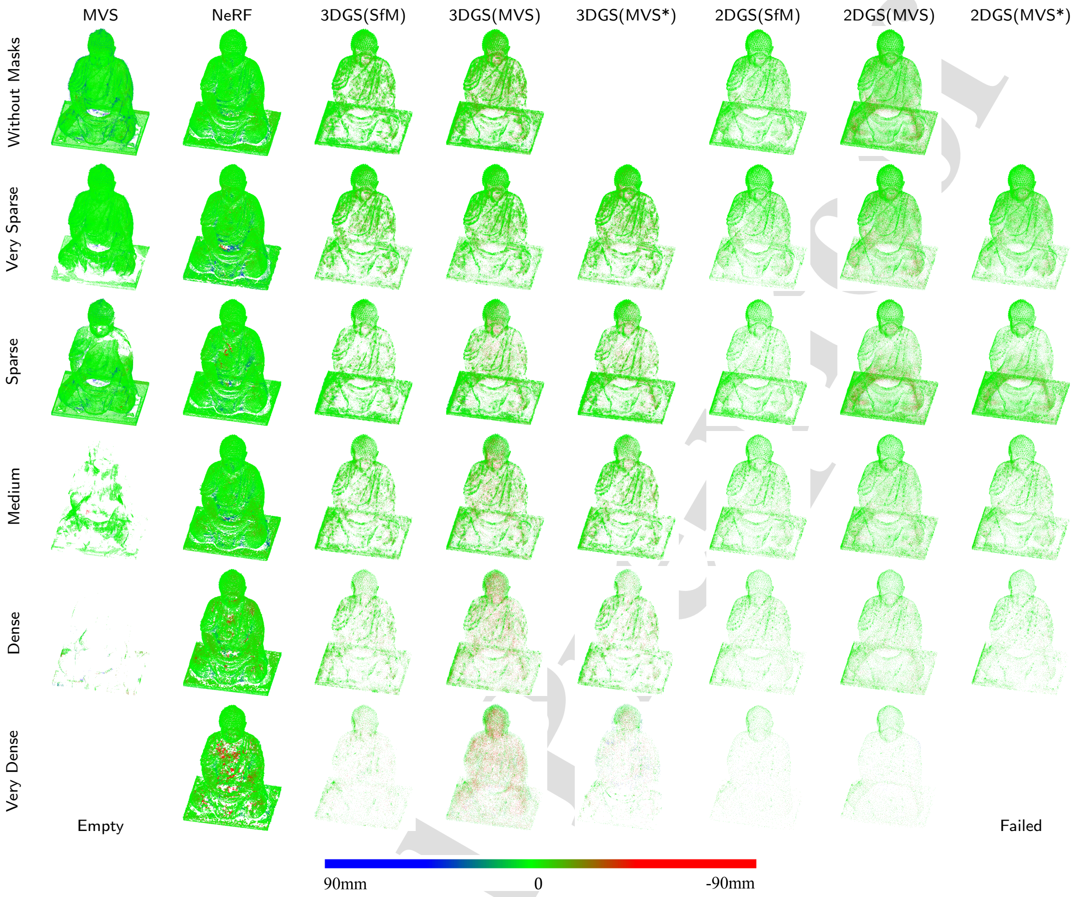


Figure 6: Geometric accuracy through cloud-to-mesh distances against the ground truth mesh for *Original*. The error displacements correspond to the color ramp. Taking into account the orientation of the normal vector the distances are signed, thus a point is considered outside the mesh when the distance is positive and inside when negative. Except MVS, the reconstructed point clouds have artifact points inside the object with high errors degrading the accuracy. MVS* depicts the MVS point cloud with corresponding masks for each occlusion level used for GS initialization. Under Very Dense masks, MVS produces only 42 background non-object points used for GS initialization causing 2DGS to fail.

the complex geometry and edges are reliably reconstructed. Subsequently, with Very Dense masks it experiences sparser reconstruction, however maintaining high completeness of 81.48% indicating its robustness. The complex geometry is reliably reconstructed due to the ability of the positional encoding to represent high frequency geometry and texture. However, the reconstruction is noisy with a substantial amount of artifact points inside the object (Figure 6) which strongly degrade the accuracy to 11.13mm. Triggered by the challenges of 3D occlusions in *Vegetation* the reconstruction is fuzzy and has noise on the outer object surface (Figure 10 and 14), but the occluded parts are more densely reconstructed compared to all other methods. With masks, the complex geometry suffers from more noise (Figure 9 and 11). The RMSE is around 9mm which is the lowest among the methods, however the point coverage is high and above 70% enough for third and second best score, without and with masks respectively (Table 5).

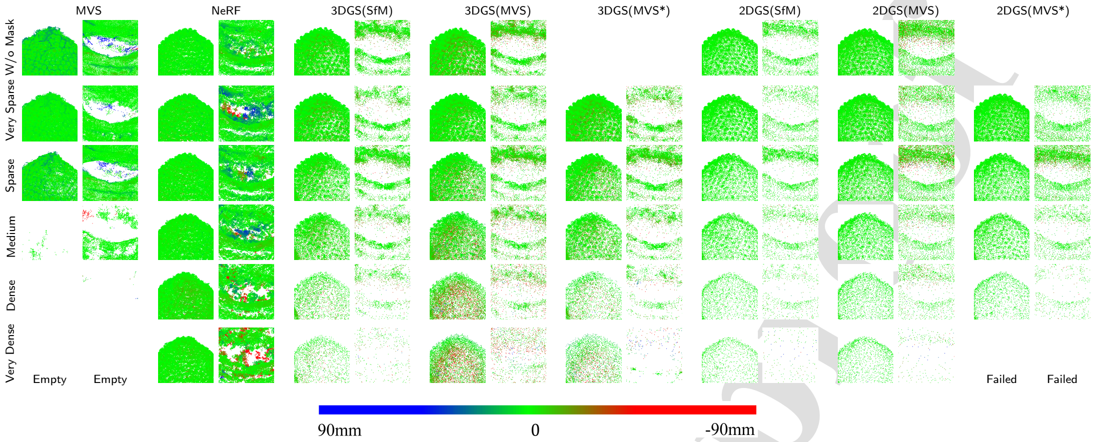


Figure 7: Cloud-to-mesh distances of complex geometry (head spikes) and homogeneous areas (lap) for *Original*. As the error displacements correspond to the color ramp, red indicates artifact points inside the object, while the blue points likely represent noise. MVS* depicts the MVS point cloud with corresponding masks for each occlusion level used for GS initialization. Under Very Dense masks, MVS produces only 42 background non-object points used for GS initialization causing 2DGS to fail.

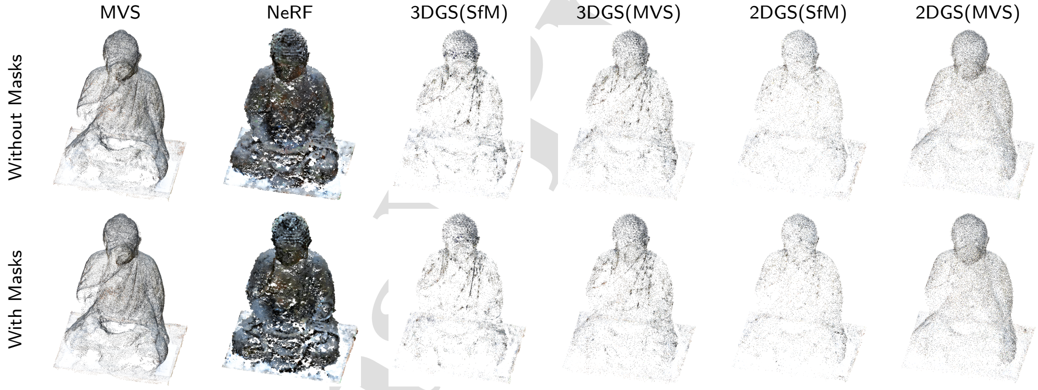


Figure 8: Point cloud reconstruction for *Vegetation* with Natural masks. MVS shows gaps in reconstructing the occluded object parts, while NeRFs achieve higher completeness and better reconstruct the underlying geometry behind occlusions. 3DGS and 2DGS completeness is conditioned by the initialization and it's higher with MVS as initial point cloud.

5.3. 3D Gaussian Splatting

In both scenarios, the densified point clouds are unstructured; complex geometry and edges are represented with a significant number of points, while the non-textured areas are sparser (Figure 4 and 8). When initializing 3DGS on MVS without and MVS* with corresponding masks for each occlusion level, the homogeneous areas are better reconstructed since MVS has more points for these parts compared to SfM. Logically, with MVS* for initialization the accuracy and completeness scores are between SfM and MVS without masks which corresponds to the number of points used for initialization, except for Dense and Very Dense masks. Nevertheless, the lap is sparse even with MVS because the point cloud also doesn't have points there (Figure 5). For *Original* without masks the results are moderate with RMSE of 4.74mm and 5.25mm (Figure 6), while the completeness is above 80% and lowest among the methods (Table 4). With only 10% occlusions, the accuracy is almost unaffected, however the point coverage slightly drops to around 80%. Increasing the occlusions to 30% and further to 50% leads to proportional drop in the RMSE

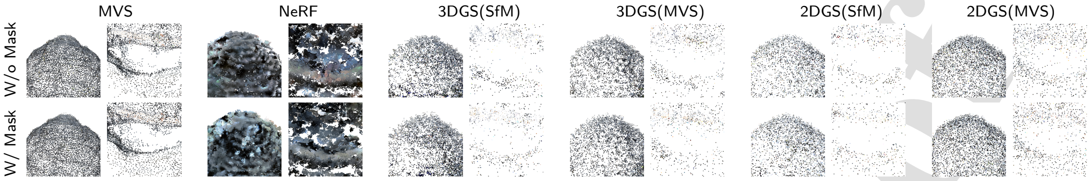


Figure 9: Enlargements of complex geometry (head spikes) and homogeneous areas (lap) for *Vegetation*. MVS struggles with the complex geometry and sparsely reconstructs the occluded homogeneous parts. NeRF better captures the head spikes and more densely the lap, however the reconstruction is very noisy. In 3DGS and 2DGS the complex geometry is already reliably reconstructed with SfM initialization, however MVS populates the homogeneous areas, increasing the completeness.

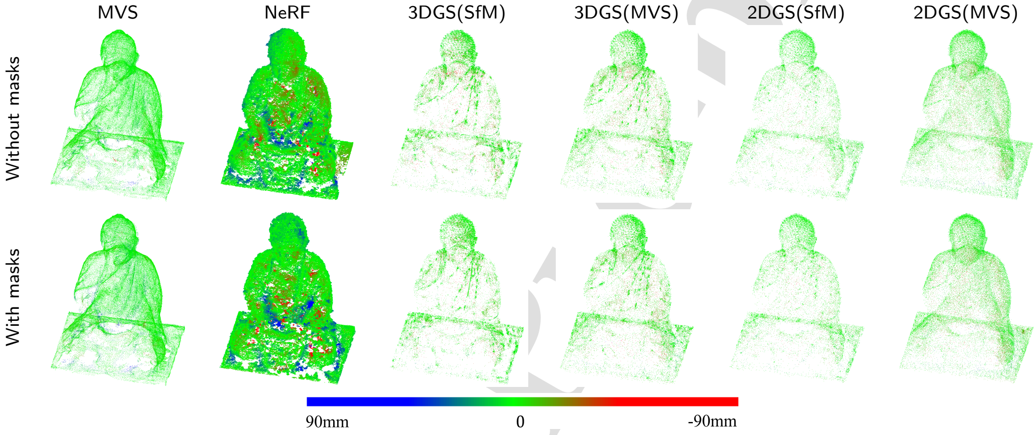


Figure 10: Geometric accuracy through cloud-to-mesh distances against the ground truth mesh for *Vegetation*. The error displacements correspond to the color ramp. Taking into account the orientation of the normal vector the distances are signed; a point is considered outside the mesh when the distance is positive and inside when negative.

to around 6mm and moderate completeness percentage below 80% regardless of the initialization. The accuracy and completeness continue to decrease as the occlusions increase (Figure 12 and 13) indicating the importance of image information as the method optimizes on full images. Under Dense masks, the RMSE subsequently drops to above 7mm, however the number of points is twice as lower than previously. When MVS* is used as initial point cloud, 3DGS achieves slightly lower completeness than with SfM, as MVS* has 10.541 points and the object is undistinguishable with only some edges reconstructed, lower than the SfM point cloud which has 42.580 points (Table 3). Nevertheless, 3DGS is able to reliably reconstruct the object with 53.54% point coverage. With Very Dense occlusions and SfM initialization, the object features are poorly reconstructed and the completeness is low 33.91%. However, using MVS without masks for initialization achieves second best completeness with relatively high 59.29%. Interestingly, when initialized on MVS* with only 42 points scattered in the scene background, 3DGS is able to reconstruct the object with more points than 3DGS(SfM). However, the object reconstruction is very noisy with a substantial amount of artifacts (Figure 7). Therefore, the RMSE reaches highest 14.16mm and the completeness lowest 23.38% among the methods. For *Vegetation*, where the occlusions are present in the training images, 3DGS(SfM) exhibits third best accuracy (Figure 10) with moderate RMSE of slightly above 7mm (Table 5). The spikes on the head are more realistically reconstructed than MVS (Figure 9) and less noisy than NeRF (Figure 11). However, it suffers from low completeness of just above 40% and 50% for SfM and MVS initialization accordingly, making it sensitive to occlusions.

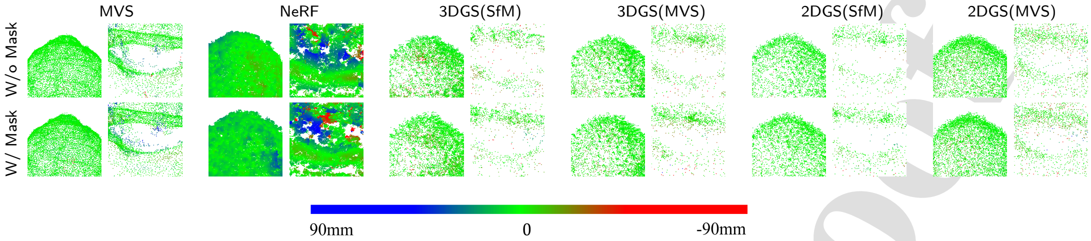


Figure 11: Cloud-to-mesh distances of complex geometry (head spikes) and homogeneous areas (lap) for *Vegetation*. As the error displacements correspond to the color ramp, red indicates artifact points inside the object, while the blue points likely represent noise.

5.4. 2D Gaussian Splatting

2DGS maintains stable performance with similar error values (Figure 6 and 10) achieving second best accuracy (Table 4 and 5), indicating robustness in the reconstruction outperforming NeRF and 3DGS. The densified point clouds have more even point distribution and better capture the object surface compared to 3DGS in both scenarios (Figure 4 and 8). 2DGS has less points than 3DGS, but higher accuracy and completeness (Figure 12 and 13) because the points are evenly spread and tightly aligned with the object surface (Figure 14) achieved by flattening the 3D Gaussians to 2D surfels and reinforcing the geometry with depth-normal regularization. Similar to 3DGS, when MVS point cloud without masks and MVS* with corresponding masks for each occlusion level is used for initialization, the completeness is higher compared to SfM with more points for homogeneous areas (Figure 5 and 9), except under Dense and Very Dense masks due to the high point sparsity of the MVS*. In *Original* without masks, 2DGS(SfM) achieves competitive accuracy results with RMSE of 3.79mm just behind NeRF and high completeness of 97.15% almost as good as MVS. The Very Sparse masks don't affect the reconstruction significantly as the accuracy and completeness are almost the same as previously. With Sparse, Medium and Dense masks the same trend is kept, 2DGS(SfM) accuracy is stable and drops insignificantly to 3.69mm, 3.89mm and 4.11mm and the completeness is still high 94.66%, 91.49% and 80.37% respectively using MVS for initialization. When 2DGS is initialized on MVS* the evaluation metrics are in between 2DGS(SfM) and 2DGS(MVS) for each occlusion level. However, when the occlusions are high 70%, 2DGS(MVS*) has less points than 2DGS(SfM) and coverage of below 60% as a consequence of the MVS sensitivity to occlusions. However, considering that MVS* has 10.541 points (Table 3) and the object features are unrecognizable with only some edges reconstructed (5.407 points), 2DGS is able to fully reconstruct the object with 59.28% point coverage indicating a consistent performance across different occlusion scenarios. Under Very Dense masks, 2DGS achieves highest correspondence with the ground truth. Nonetheless, it experiences a sharp decline in completeness to 29.14% with SfM and 40.98% with MVS for initialization, enough for third best results. While maintaining high accuracy, only the object shape can be distinguished and the geometric features are poorly reconstructed. Challenged by the sparsity of the initial point cloud which has only 42 scattered background points not belonging to the object, 2DGS(MVS*) fails to reconstruct the object. After around 11.000 iterations it runs Out Of Memory (OOM) caused by an explosion in Gaussian count triggered by elongated (stretched in one direction) Gaussians. Although it experiences a decrease in accuracy and completeness in *Vegetation* due to the gaps in the geometry of the occluded parts, 2DGS(SfM) shows second best results with moderate RMSE within 6mm and high completeness of above 70% (Table 5), again outperforming 3DGS while being competitive with MVS and NeRF (Figure 9 and 11).

Overall, MVS performs best in both scenarios without occlusions, but the performance manifests a strong descent as the occlusion coverage increases. Although it excels in pin-point accuracy, the completeness proportionally drops until 30% masked pixels, then it shows a sharp decline with 50%, eventually failing to reconstruct the object with 90% occlusions. On the other hand, radiance field methods provide higher point coverage and perform better under severe occlusions indicating a consistent performance across different occlusion scenarios. NeRF shows robustness in the reconstruction with highest completeness considering occlusions. In spite of the accuracy proportionally decreasing with increased mask coverage, the surface representation remains stable. 3DGS regardless of the initialization, struggles with accuracy and completeness. The point clouds are unstructured with uneven surface point distribution, guided only from the input point cloud and photometric loss. 2DGS generally shows better correspondence with the

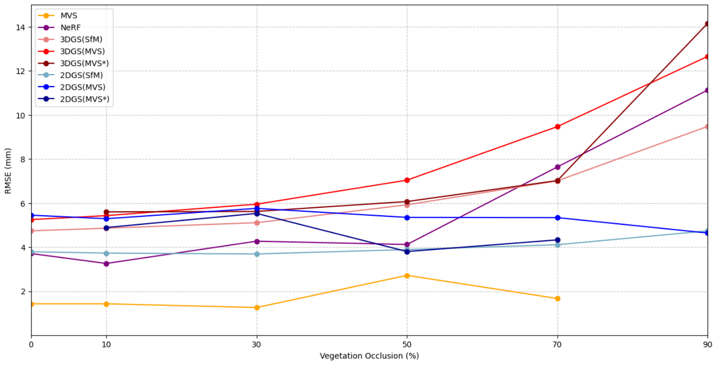


Figure 12: Geometric accuracy for *Original* with Synthetic masks for each 3D reconstruction method through RMSE. The correlation has a linear trend; as the vegetation occlusion percentage increases the accuracy proportionally decreases.

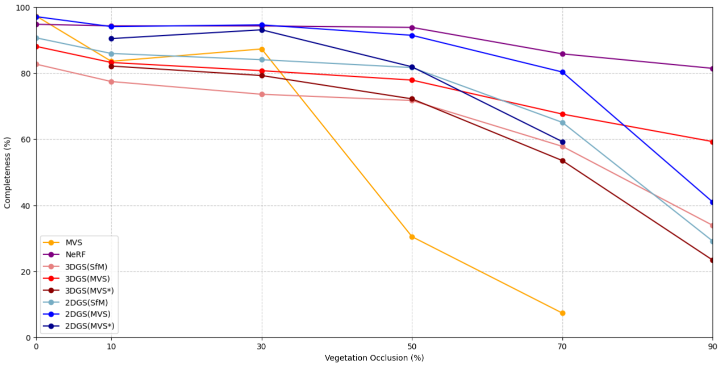


Figure 13: Object surface coverage for *Original* under Synthetic occlusions for each 3D reconstruction method. Increasing the occlusion percentage proportionally degrades the completeness for the radiance field methods. However, MVS shows a sharp decline when the occlusions are above 50%, failing to reconstruct the object when 90%.

ground truth and higher completeness compared to 3DGS, representing object's geometry through the median of 2D Gaussian disks, additionally enforced with depth and normal regularization. The depth constraint prevents floating Gaussians at different depth in different views. Additionally, inferring surface normals from the depth maps enforces normal consistency across views preventing discontinuities ensuring that Gaussians from different view-points obey the same surface orientation. Hence, 2DGS(SfM) exhibits second best accuracy results outperforming NeRF in both scenarios, while 2DGS(MVS) achieves high completeness without significantly sacrificing the accuracy, due to the more densely reconstructed homogeneous areas. For both GS methods, when MVS* with corresponding masks for each occlusion level is used for initialization, the accuracy and completeness are between SfM and MVS without masks which is logical due to the number of points in the initial point cloud. This holds until Dense and Very Dense masks, where the point coverage is slightly lower than with SfM initialization, as a consequence to MVS high sensitivity to occlusions. NeRF, 3DGS and 2DGS show lower accuracy than MVS since the optimization is based on minimizing the difference between predicted and actual pixel color from the input images through gradient descent without additional geometric constraints. Moreover, artifact points inside the object which are projections from input views that weren't moved into their correct place geometrically are present which distort the accuracy and don't contribute to the completeness. Thus, the error displacements tend to increase faster to the negative because those are the points behind the mesh surface. The errors with positive values lie above the nearest mesh triangle and thus most likely represent noise and outlier points. We can also observe color differences in the object reconstruction between the scenarios. *Original* is captured indoors under controlled lighting, which is different from *Vegetation* captured outdoors under natural illumination. However, apart from visual appearance, the color differences don't affect the geometry.

6. Discussion

Although MVS shows almost a uniform match with the ground truth (Figure 6 and 10), it sparsely reconstructs homogeneous areas and is unable to reconstruct the geometry behind occlusions (Figure 4 and 8). The surface point coverage experiences a sharp decline under Medium occlusions (Figure 13) falling short to reconstruct the object with Very Dense masks because the object needs to be visible in a significant number of images for a complete reconstruction, making MVS sensitive to occlusions.

Despite the Very Dense masks covering 90% of the images almost entirely occluding the object, NeRF can still infer the underlying geometry and appearance with high surface coverage of 81.48% (Table 4). We argue that since we apply the same Synthetic mask to all images, the non-masked image pixels are well distributed across the images captured on a hemispheric trajectory completely covering the object from all sides. Thus, even with reduced image information, the accumulated density and color from different view-points through the visible pixels allow NeRF to reconstruct a coherent geometry. The high-dimensional positional encoding enables the MLP to learn spatially smooth representation. However, it might hallucinate as a consequence of the restricted image information, containing a substantial number of artifact points inside the object (Figure 7) which explains the low accuracy of above 11mm. Moreover, as the density field is given implicitly, we extract the geometry as a point cloud using voxelization which enforces spatial consistency. We then apply a 3D density-gradient based Canny edge detection filter. The density gradients are determined independently of the absolute magnitude of density values, enabling extraction of edges in regions with lower density within the field, contributing to the completeness.

NeRF, 3DGS and 2DGS reconstructions in both scenarios show lower accuracy than MVS due to the lack of geometric constraints during optimization, minimizing the photometric error between rendered and ground truth images. The artifact points inside the object additionally degrade the accuracy. However, all radiance field methods show higher point coverage under occlusions as they are able to approximate the geometry of the occluded object parts. In addition, the completeness in 3DGS and 2DGS depends on the initial point cloud. We use SfM and MVS without masks and MVS* with corresponding masks for each occlusion level accordingly. With SfM initialization, the completeness is moderate due to the sparsely reconstructed homogeneous object parts (Figure 5) caused by the lack of points for these parts in the SfM point cloud, since it requires visually distinctive points for feature extraction. The point coverage increases when initialized on MVS point cloud especially under Dense and Very Dense Synthetic masks as well as with Natural masks (Table 5). It is worth noting that even though the completeness increases, the accuracy slightly decreases (Figure 12) as more points are present inside the object which were not placed to their correct geometric position. However, the number of points in the densified GS point clouds is smaller than the number of points of MVS used for initialization. We argue that because of the high number of points, the Gaussians were too small and got merged. During optimization the Gaussians shrink and grow, but are also periodically pruned based on opacity and size constraints. Points that are either too transparent, too large in screen space or too large in world space are removed. When MVS* is used, logically the results are between SfM and MVS without masks and the completeness is higher than with SfM, until Dense masks when 3DGS(MVS*) and 2DGS(MVS*) exhibit slightly lower point coverage than when initialized on SfM due to the point sparsity. However, the scores are still relatively high considering that the MVS* containing 10.541 points (Table 3), much less than SfM that has 42.580 points and the masks are also applied on the training images. 3DGS and 2DGS are able to approximate the object parts affected by occlusions even with very sparse initialization and Dense masks (70%) applied on the images, which indicates their strength and robustness under occlusions.

Interestingly, 3DGS can reconstruct the object even when initialized on MVS* with Very Dense masks with only 42 points (none of which belonging to the object) and trained with 90% masked image information. The algorithm optimizes the 3D Gaussian primitives in 3D space interpolating in all three dimensions, filling in the gaps and refining the geometry over training iterations. Although the densified point cloud has 14.217 points which is higher compared to 3DGS(SfM), the completeness is relatively low 23.38%. The reconstruction is very noisy, the object structure and features are hard to distinguish, thus the RMSE is above 14mm and lowest among the methods. However, considering that the object didn't have any points before training, the results are satisfactory. Without affecting the geometry, we can notice that the point cloud is not color coded. As initialized on only 42 background points not covering enough of the scene and not well distributed, the optimization leads to black splats because the algorithm struggles to infer proper color. This arises due to insufficient color constraints since the color information is too sparse for optimization. Gaussians initialized in empty regions may have no meaningful color supervision and moreover, there aren't enough color constraints from the input images either as the masks cover 90% of the image pixels. In contrast, 2DGS fails to reconstruct a meaningful densified point cloud. Initialized on only 42 non-object points and trained with 10% image pixels, the algorithm doesn't have a strong depth prior and therefore is unable to establish a coherent scene representation. Since the normals depend on depth gradients across views and 2DGS lacks explicit depth constraints, achieving normal consistency is challenging. Instead, per-view Gaussians may be misaligned, leading to incorrect surface orientation. Depth is implicitly enforced if more points capture the scene geometry, while normals are better estimated when more points define the geometry. Thus, 2D Gaussians with extreme elongation (stretched in one direction) can collapse to very small points in screen space. With high opacity, their movement can cause significant pixel changes, leading to pronounced positional gradients. Moreover, some points project smaller than one pixel, resulting in their covariance being replaced by a fixed value through the antialiased low-pass filter. Consequently, these points cannot properly adjust their scaling and rotation causing rapid gradient accumulation (Liu et al., 2024). This triggers exponential increase in Gaussian count because 2DGS keeps adding Gaussians per view (almost 2M Gaussians until 7.000 iterations), leading to noise and ultimately OOM after approximately 11.000 iterations.

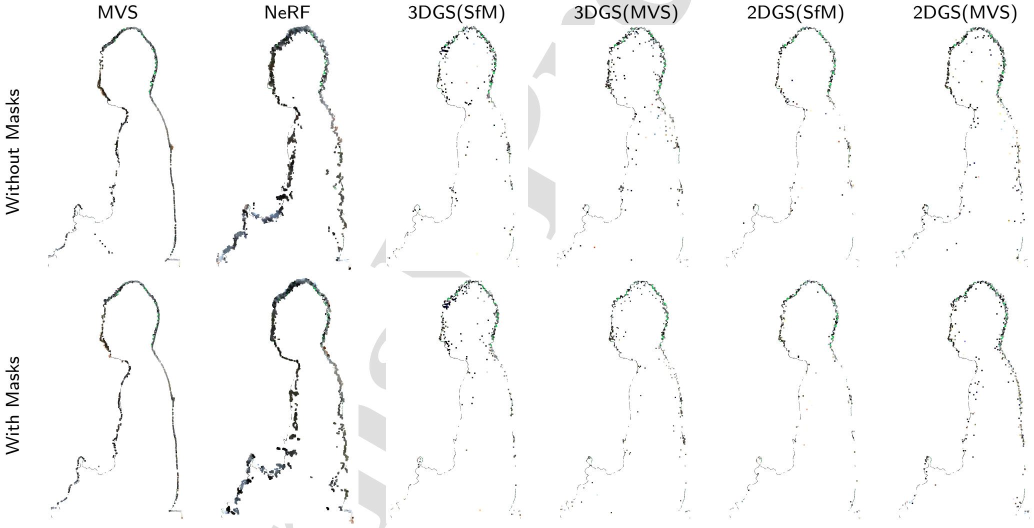


Figure 14: Profiles of the reconstructed point clouds against the ground truth (black thin line) for *Vegetation*. MVS benefits from the occlusion masks eliminating the artifact points inside the object. However, the masks have little impact on the radiance field methods. NeRF shows noisy reconstruction and slightly higher completeness. Compared to 3DGS, 2DGS better represents object surface with less scattered points. When MVS is used for initialization, more points are visible on the object outer surface as well as inside. The points are 6x increased for visualization.

However, the number of points doesn't necessarily imply higher completeness. MVS with Very Sparse masks has more points (739.936) than with Sparse masks (554.636), but a slightly lower completeness of 83.62% and 87.36% respectively. The gaps in the reconstruction correspond to the masked pixels. The Very Sparse mask are concentrated on the lower object part and plate and the grass is in form of a cluster with no empty spaces in between, while the Sparse mask consists of branches with empty spaces between the leaves distributed equally on the whole image (Table

1). This means that the mask coverage and placement affect MVS reconstruction, unlike radiance field methods in which the number of points proportionally depends on the mask percentage. In all NeRF point clouds the number of points is significantly higher among the methods, but the completeness doesn't linearly follow this trend because they are not evenly distributed on the object outer surface.

Reasonably, masking out regions results in free-occlusion reconstruction (Figure 3), but it degrades the performance across all methods. In NeRF the density for the masked parts is zero, while 3DGS and 2DGS don't have enough information to adjust the Gaussians to match the input images. Hence, the images have bigger influence on the results than the point cloud used for initialization as we used the same SfM and MVS without masks as input in all occlusion mask variations and the completeness drops with higher occlusion percentage. However, we only considered the masks in the training as we are focused on their influence on the geometric reconstruction. To analyze how the vegetation occlusions affect the pose estimation, the masks should be considered in the poses needed for training NeRF and subsequently in the sparse SfM point cloud used for initialization in GS methods.

In spite of that, the Natural masks seem to have a low effect on the results of the radiance field methods, especially GS. In *Vegetation* where the occlusions exist in the scene, we can notice scattered points at the bottom of the object, which are eliminated considering masks improving MVS accuracy (Figure 14). NeRF achieves slightly higher completeness with masks (Figure 9) but lowest accuracy (Figure 11), while 3DGS and 2DGS with MVS initialization strike the best balance between accuracy and completeness and show robustness in the reconstruction. We didn't use MVS* with Natural masks for GS initialization because the reconstructed point clouds without and with masks have similar accuracy and completeness scores with almost identical number of points. The vegetation acts as a foreground occlusion on just 50 out of 125 images with relatively sparse foliage of quite low 35% (Figure 2) and doesn't seem to impact the geometric reconstruction. In future work a higher occlusion percentage persistent on more images should be considered.

7. Conclusion

In summary, we provide a comprehensive qualitative and quantitative analysis of the accuracy and completeness of the 3D geometry behind vegetation occlusions reconstructed by traditional MVS and radiance field methods namely: NeRF, 3DGS and 2DGS in real-world scenarios. To investigate which method can provide most reliable results with least image information, we consider Synthetic masks with different occlusion coverage starting from 10% to 90% masked pixels with 20% increment. The same synthetic mask is applied to all images of an occlusion-free indoor scenario to have control over the mask percentage and ensure it remains constant on all images. Additionally, we consider two initialization strategies in 3DGS and 2DGS, namely SfM and MVS without masks and MVS* with corresponding masks for each occlusion level to gather more insights how the initial point cloud affects the overall reconstruction. In order to assess the effect of spatially consistent 3D occlusions, we use Natural masks where the vegetation is stationary in the 3D scene, but relative to the view-point. Hence, only the foreground vegetation is considered as occlusion with an average of 35% masked pixels. The key challenge lies in recovering the underlying geometry behind occlusion, thus investigating if radiance field methods can compete against traditional MVS for scenarios where the latter falls short.

Generally, MVS shows lowest accuracy errors, however the completeness manifests a sharp decline as the occlusion percentage increases, eventually failing to reconstruct the object with 90% masks. NeRF exhibits robustness in the reconstruction with highest completeness considering masks, although the accuracy proportionally decreases with higher mask occlusion percentage. 3DGS, regardless of the initialization, struggles with accuracy and completeness in both scenarios. 2DGS shows better correspondence with the ground truth and higher completeness compared to 3DGS. 2DGS(SfM) achieves second best accuracy results right behind MVS, outperforming NeRF in both scenarios, while 2DGS(MVS) shows high completeness without significantly sacrificing the accuracy, due to the more densely reconstructed homogeneous areas. We demonstrate that radiance field methods can compete against traditional MVS, showing robust performance for a complete reconstruction under severe vegetation occlusions.

In future research, the influence of the vegetation occlusions on the poses and subsequently on the sparse SfM point cloud used for GS initialization and how that reflects on the quality of the 3D reconstruction should be investigated. Moreover, the study should be extended to spatially consistent 3D occlusions that are stationary and exist in the scene from which Natural masks can be derived, taking into account different vegetation types and higher vegetation coverage present in all images.

References

Barron, J.T., Mildenhall, B., Tancik, M., Hedman, P., Martin-Brualla, R., Srinivasan, P.P., 2021. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 5855–5864.

Barron, J.T., Mildenhall, B., Verbin, D., Srinivasan, P.P., Hedman, P., 2022. Mip-nerf 360: Unbounded anti-aliased neural radiance fields, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5470–5479.

Besl, P.J., McKay, N.D., 1992. Method for registration of 3-d shapes, in: Sensor fusion IV: control paradigms and data structures, Spie. pp. 586–606.

Chen, H., Li, C., Lee, G.H., 2023. Neusg: Neural implicit surface reconstruction with 3d gaussian splatting guidance. arXiv preprint arXiv:2312.00846 .

Chen, X., Zhang, Q., Li, X., Chen, Y., Feng, Y., Wang, X., Wang, J., 2022. Hallucinated neural radiance fields in the wild, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12943–12952.

Dai, P., Xu, J., Xie, W., Liu, X., Wang, H., Xu, W., 2024. High-quality surface reconstruction using gaussian surfels, in: ACM SIGGRAPH 2024 Conference Papers, pp. 1–11.

Foroutan, Y., Rebain, D., Yi, K.M., Tagliasacchi, A., 2024. Does gaussian splatting need sfm initialization? arXiv preprint arXiv:2404.12547 .

Fridovich-Keil, S., Yu, A., Tancik, M., Chen, Q., Recht, B., Kanazawa, A., 2022. Plenoxels: Radiance fields without neural networks, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 5501–5510.

Fu, Y., Liu, S., Kulkarni, A., Kautz, J., Efros, A.A., Wang, X., 2023. Colmap-free 3d gaussian splatting arXiv:2312.07504.

Guédon, A., Lepetit, V., 2024. Sugar: Surface-aligned gaussian splatting for efficient 3d mesh reconstruction and high-quality mesh rendering, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5354–5363.

Huang, B., Yu, Z., Chen, A., Geiger, A., Gao, S., 2024. 2d gaussian splatting for geometrically accurate radiance fields, in: ACM SIGGRAPH 2024 Conference Papers, pp. 1–11.

Jäger, M., Hübner, P., Haitz, D., Jutzi, B., 2023. A comparative neural radiance field (nerf) 3d analysis of camera poses from hololens trajectories and structure from motion. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLVIII-1/W1-2023, 207–213. doi:10.5194/isprs-archives-XLVIII-1-W1-2023-207-2023.

Jäger, M., Jutzi, B., 2023. 3d density-gradient based edge detection on neural radiance fields (nerfs) for geometric reconstruction. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 48, 71–78.

Karami, A., Rigon, S., Mazzacca, G., Yan, Z., Remondino, F., 2023. Nerfbk: A high-quality benchmark for nerf-based 3d reconstruction. arXiv preprint arXiv:2306.06300 .

Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G., 2023. 3d gaussian splatting for real-time radiance field rendering. ACM Trans. Graph. 42, 139–1.

Kheradmand, S., Rebain, D., Sharma, G., Sun, W., Tseng, J., Isack, H., Kar, A., Tagliasacchi, A., Yi, K.M., 2024. 3d gaussian splatting as markov chain monte carlo. arXiv preprint arXiv:2404.09591 .

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al., 2023. Segment anything, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4015–4026.

Kulhanek, J., Peng, S., Kukelova, Z., Pollefeys, M., Sattler, T., 2024. Wildgaussians: 3d gaussian splatting in the wild. arXiv preprint arXiv:2407.08447 .

Lee, J., Kim, I., Heo, H., Kim, H.J., 2023. Semantic-aware occlusion filtering neural radiance fields in the wild. arXiv preprint arXiv:2303.03966 .

Li, H., Liu, J., Sznaier, M., Camps, O., 2024a. 3d-hgs: 3d half-gaussian splatting. arXiv preprint arXiv:2406.02720 .

Li, Y., Lyu, C., Di, Y., Zhai, G., Lee, G.H., Tombari, F., 2024b. Geogaussian: Geometry-aware gaussian splatting for scene rendering. arXiv preprint arXiv:2403.11324 .

Li, Z., Müller, T., Evans, A., Taylor, R.H., Unberath, M., Liu, M.Y., Lin, C.H., 2023. Neuralangelo: High-fidelity neural surface reconstruction, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8456–8465.

Liu, Y., Luo, C., Mao, Z., Peng, J., Zhang, Z., 2024. Citygaussianv2: Efficient and geometrically accurate reconstruction for large-scale scenes. arXiv preprint arXiv:2411.00771 .

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. International journal of computer vision 60, 91–110.

Martin-Brualla, R., Radwan, N., Sajjadi, M.S., Barron, J.T., Dosovitskiy, A., Duckworth, D., 2021. Nerf in the wild: Neural radiance fields for unconstrained photo collections, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7210–7219.

Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R., 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM 65, 99–106.

Müller, T., Evans, A., Schied, C., Keller, A., 2022. Instant neural graphics primitives with a multiresolution hash encoding. ACM Transactions on Graphics (ToG) 41, 1–15.

Oechsle, M., Peng, S., Geiger, A., 2021. Unisurf: Unifying neural implicit surfaces and radiance fields for multi-view reconstruction, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 5589–5599.

Petrovska, I., Jäger, M., Haitz, D., Jutzi, B., 2023. Geometric accuracy analysis between neural radiance fields (nerfs) and terrestrial laser scanning (tls). The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 48, 153–159.

Petrovska, I., Jutzi, B., 2024. Vision through obstacles-3d geometric reconstruction and evaluation of neural radiance fields (nerfs). Remote Sensing 16, 1188.

Sabour, S., Goli, L., Kopanas, G., Matthews, M., Lagun, D., Guibas, L., Jacobson, A., Fleet, D.J., Tagliasacchi, A., 2024. Spotlesssplats: Ignoring distractors in 3d gaussian splatting. arXiv preprint arXiv:2406.20055 .

Sabour, S., Vora, S., Duckworth, D., Krasin, I., Fleet, D.J., Tagliasacchi, A., 2023. Robustnerf: Ignoring distractors with robust losses, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 20626–20636.

Schonberger, J.L., Frahm, J.M., 2016. Structure-from-motion revisited, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4104–4113.

Schönberger, J.L., Zheng, E., Frahm, J.M., Pollefeys, M., 2016. Pixelwise view selection for unstructured multi-view stereo, in: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14, Springer. pp. 501–518.

Sitzmann, V., Zollhöfer, M., Wetzstein, G., 2019. Scene representation networks: Continuous 3d-structure-aware neural scene representations. Advances in Neural Information Processing Systems 32.

Sun, C., Sun, M., Chen, H.T., 2022. Improved direct voxel grid optimization for radiance fields reconstruction. arXiv preprint arXiv:2206.05085 .

Sun, W., Trulls, E., Tseng, Y.C., Sambandam, S., Sharma, G., Tagliasacchi, A., Yi, K.M., 2023. Pointnerf++: A multi-scale, point-based neural radiance field. arXiv preprint arXiv:2312.02362 .

Wei, P., Yan, L., Xie, H., Qiu, D., Qiu, C., Wu, H., Zhao, Y., Hu, X., Huang, M., 2024. Lidenerf: Neural radiance field reconstruction with depth prior provided by lidar point cloud. ISPRS Journal of Photogrammetry and Remote Sensing 208, 296–307.

Wolf, Y., Bracha, A., Kimmel, R., 2024. Gs2mesh: Surface reconstruction from gaussian splatting via novel stereo views, in: ECCV 2024 Workshop on Wild 3D: 3D Modeling, Reconstruction, and Generation in the Wild.

Xu, C., Kerr, J., Kanazawa, A., 2024a. Splatfacto-w: A nerfstudio implementation of gaussian splatting for unconstrained photo collections. arXiv preprint arXiv:2407.12306 .

Xu, Q., Xu, Z., Philip, J., Bi, S., Shu, Z., Sunkavalli, K., Neumann, U., 2022. Point-nerf: Point-based neural radiance fields, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5438–5448.

Xu, W., Gao, H., Shen, S., Peng, R., Jiao, J., Wang, R., 2024b. Mvpgs: Excavating multi-view priors for gaussian splatting from sparse input views, in: European Conference on Computer Vision, Springer. pp. 203–220.

Zhu, C., Wan, R., Tang, Y., Shi, B., 2023. Occlusion-free scene recovery via neural radiance fields, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 20722–20731.

Zimny, D., Trzciński, T., Spurek, P., 2022. Points2nerf: Generating neural radiance fields from 3d point cloud. arXiv preprint arXiv:2206.01290 .

Table 4

Quantitative results of the cloud-to-mesh comparison addressing accuracy and completeness for *Original* under Synthetic masks. 3DGS and 2DGS are initialized on SfM and MVS without and MVS* with corresponding masks for each occlusion level. MVS excels in pin-point accuracy, but fails to reconstruct the object under Very Dense masks producing only 42 background scene points used for GS initialization. 2DGS(SfM) achieves second best accuracy results, while NeRF shows highest completeness. The first, second and third best results are highlighted.

Scenario (*Original*)	Method	Accuracy (mm)			Completeness (%)	
		Mean ↓	SD ↓	RMSE ↓	Npts	Cpl ↑
Without Masks	MVS	0.16	1.42	1.43	845.456	97.35
	NeRF	-0.58	3.66	3.71	938.178	94.83
	3DGS(SfM)	-2.02	4.29	4.74	352.298	82.75
	3DGS(MVS)	-2.76	4.47	5.25	377.291	88.14
	2DGS(SfM)	-1.22	3.59	3.79	156.465	90.73
	2DGS(MVS)	-2.25	4.97	5.45	235.238	97.15
Very Sparse Masks	MVS	0.14	1.42	1.43	739.936	83.62
	NeRF	-0.36	3.24	3.26	937.638	94.35
	3DGS(SfM)	-2.23	4.31	4.86	251.892	77.49
	3DGS(MVS)	-2.57	4.79	5.43	269.526	83.26
	3DGS(MVS*)	-3.09	4.67	5.60	264.594	82.18
	2DGS(SfM)	-1.08	3.57	3.73	113.126	85.99
	2DGS(MVS)	-1.84	4.96	5.29	160.881	94.13
	2DGS(MVS*)	-1.58	4.63	4.89	151.550	90.50
Sparse Masks	MVS	0.05	1.26	1.26	554.636	87.36
	NeRF	-0.55	4.23	4.27	910.699	94.33
	3DGS(SfM)	-2.14	4.64	5.11	201.356	73.63
	3DGS(MVS)	-2.65	5.33	5.95	217.345	80.78
	3DGS(MVS*)	-2.80	4.88	5.62	206.068	79.31
	2DGS(SfM)	-1.06	3.53	3.69	92.719	84.11
	2DGS(MVS)	-2.40	5.23	5.76	154.409	94.66
	2DGS(MVS*)	-2.26	5.04	5.53	146.057	93.16
Medium Masks	MVS	-0.02	2.72	2.72	73.937	30.49
	NeRF	-0.81	4.04	4.12	908.031	93.89
	3DGS(SfM)	-3.04	5.09	5.92	128.607	71.75
	3DGS(MVS)	-3.46	6.13	7.04	143.167	77.93
	3DGS(MVS*)	-2.99	5.28	6.07	127.026	72.24
	2DGS(SfM)	-1.21	3.70	3.89	62.587	81.72
	2DGS(MVS)	-1.81	5.04	5.35	87.876	91.49
	2DGS(MVS*)	-1.24	3.59	3.80	65.277	81.95
Dense Masks	MVS	-0.01	1.67	1.67	5.407	7.34
	NeRF	-2.60	7.18	7.64	861.056	85.87
	3DGS(SfM)	-3.39	6.14	7.01	50.505	57.85
	3DGS(MVS)	-5.13	7.96	9.47	68.565	67.63
	3DGS(MVS*)	-2.94	6.38	7.02	49.090	53.54
	2DGS(SfM)	-1.01	3.99	4.11	27.676	65.11
	2DGS(MVS)	-1.75	5.04	5.34	42.677	80.37
	2DGS(MVS*)	-0.76	4.26	4.33	25.020	59.28
Very Dense Masks	MVS	-	-	-	0	0
	NeRF	-3.66	10.51	11.13	778.341	81.48
	3DGS(SfM)	-4.08	8.56	9.48	13.215	33.91
	3DGS(MVS)	-7.34	10.31	12.66	44.459	59.29
	3DGS(MVS*)	-2.44	13.94	14.16	14.217	23.38
	2DGS(SfM)	-0.46	4.72	4.74	6.198	29.14
	2DGS(MVS)	-0.49	4.62	4.64	9.728	40.98
	2DGS(MVS*)	Failed				

Table 5
Quantitative results of the cloud-to-mesh comparison addressing the accuracy and completeness for *Vegetation* under Natural masks. 3DGS and 2DGS are initialized on SfM and MVS point clouds without masks. 2DGS(SfM) exhibits second best accuracy results, just behind MVS. The first, second and third best results are highlighted.

Scenario	Method	Accuracy (mm)			Completeness (%)	
(*Vegetation*)		Mean ↓	SD ↓	RMSE ↓	Npts	Cpl ↑
Without Masks	MVS	0.53	3.18	3.23	117.427	75.95
	NeRF	-0.16	9.06	9.06	989.902	70.78
	3DGS(SfM)	-2.54	7.05	7.49	40.807	42.71
	3DGS(MVS)	-3.28	7.99	8.64	42.105	51.90
	2DGS(SfM)	-0.86	6.07	6.13	23.248	49.17
	2DGS(MVS)	-2.38	7.26	7.63	44.496	71.94
With Masks	MVS	0.61	2.71	2.77	117.428	78.19
	NeRF	0.09	9.38	9.38	1.006.648	75.26
	3DGS(SfM)	-2.48	6.82	7.26	41.722	43.35
	3DGS(MVS)	-2.71	7.88	8.34	43.166	52.90
	2DGS(SfM)	-1.42	6.08	6.24	23.865	48.98
	2DGS(MVS)	-2.43	7.17	7.57	45.907	70.39

Declaration of Interest Statement

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The author is an Editorial Board Member/Editor-in-Chief/Associate Editor/Guest Editor for this journal and was not involved in the editorial review or the decision to publish this article.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: