*remote sensing*

MDPI

*Article*

# Vision through Obstacles—3D Geometric Reconstruction and Evaluation of Neural Radiance Fields (NeRFs)

**Ivana Petrovska * and Boris Jutzi**

Institute of Photogrammetry and Remote Sensing (IPF), Karlsruhe Institute of Technology (KIT), 76131 Karlsruhe, Germany; boris.jutzi@kit.edu

* Correspondence: ivana.petrovska@partner.kit.edu

**Abstract:** In this contribution we evaluate the 3D geometry reconstructed by Neural Radiance Fields (NeRFs) of an object's occluded parts behind obstacles through a point cloud comparison in 3D space against traditional Multi-View Stereo (MVS), addressing the accuracy and completeness. The key challenge lies in recovering the underlying geometry, completing the occluded parts of the object and investigating if NeRFs can compete against traditional MVS for scenarios where the latter falls short. In addition, we introduce a new "obSTaclE, occLusion and visibiLity constrAints" dataset named STELLA concerning transparent and non-transparent obstacles in real-world scenarios since there is no existing dataset dedicated to this problem setting to date. Considering that the density field represents the 3D geometry of NeRFs and is solely position-dependent, we propose an effective approach for extracting the geometry in the form of a point cloud. We voxelize the whole density field and apply a 3D density-gradient based Canny edge detection filter to better represent the object's geometric features. The qualitative and quantitative results demonstrate NeRFs' ability to capture geometric details of the occluded parts in all scenarios, thus outperforming in completeness, as our voxel-based point cloud extraction approach achieves point coverage up to 93%. However, MVS remains a more accurate image-based 3D reconstruction method, deviating from the ground truth 2.26 mm and 3.36 mm for each obstacle scenario respectively.

**Keywords:** neural radiance fields; geometry evaluation; point clouds; obstacles; multi-view stereo; 3D reconstruction; new dataset



**Citation:** Petrovska, I.; Jutzi, B. Vision through Obstacles—3D Geometric Reconstruction and Evaluation of Neural Radiance Fields (NeRFs). *Remote Sens.* **2024**, *16*, 1188. https://doi.org/10.3390/rs16071188

Academic Editors: Fabio Remondino, Jiaojiao Li and Rongjun Qin

Received: 20 February 2024
Revised: 13 March 2024
Accepted: 22 March 2024
Published: 28 March 2024



## 1. Introduction

Image-based 3D reconstruction techniques uncover many applications in documenting the geometric dimensions of the environment. Geometric scene reconstruction in 3D space usually contains the depth information and relative positions between two or more entities, which makes it suitable for applications that require highly accurate positioning information. Providing a detailed and unambiguous representation of the spatial arrangement and geometry, point clouds have become a fundamental data structure [1] for realistic representation of objects and scenes in applications where high geometric accuracy and photo-realism must be satisfied. Differing from images that are arranged as a grid pattern and where the neighborhood of a pixel can easily be determined, point clouds represent scenes through a collection of non-uniform distributed points [2], where each point corresponds to a specific location or feature in the real world. The versatility of point clouds lies in the level of details for representing complex geometry, determined by the number of points.

Reliably reconstructing a scene in 3D space in the form of a dense point cloud based on overlapping images captured from different viewpoints is originally addressed by Multi-View Stereo (MVS) [3]. After the camera position and orientation have been identified through Structure-from-Motion (SfM) [4], MVS relies on cross-view correspondence matching and triangulation to estimate the pixel-wise depth value by minimizing the gap

between the predicted point and ray intersection. It can produce high-quality results when applied on well-defined textured surfaces; thus, the main challenge lies in computing distinctive points or features in the images of the scene [5]. Therefore, traditional MVS often fails to obtain an accurate geometric reconstruction, particularly in cases of a lack of texture, texture repetition, illumination changes or occlusions.

Recently, scene representation networks [6] have shown great potential to address these limitations. Neural Radiance Fields (NeRFs) [7] have revolutionized the traditional methods by storing 3D representations within the weights of a neural network from images and paired camera parameters. The scene is represented as a continuous volumetric field where the position-dependent density and radiance, which additionally depends on the viewing direction at any given 3D point, are predicted through ray tracing from the corresponding overlapping images and viewing directions. The explicit geometry of NeRFs in the form of a point cloud is gained by sampling 3D points along the rays, resulting in rendering of depth maps.

Despite offering new insights in generating novel views from images, existing image-based 3D reconstruction methods do not consider obstacles which is paramount, since a real-world scene often comprises multiple objects, with one or more being occluded by other objects. In cases where objects in the scene obstruct each other and certain parts of the object are not visible due to occlusions or limited camera angles, it can be challenging to generate a complete and accurate 3D representation as feature extraction and matching become restricted, especially for complex geometry. The quality of occluded object reconstruction is important because obstacles are common in our daily environment and furthermore, it can indicate how accurately NeRFs can represent the volumetric field under occlusions [8].

We introduce a new dataset named STELLA (available online at https://github.com/sqirrel3/STELLA (accessed on 12 January 2024)) which stands for "obSTaclE, occLusion and visibiLity constrAints", concerning transparent and non-transparent obstacles since this area remains rarely explored to date. The dataset consists of real-world challenging scenarios rather than synthetic in order to address users not only in research but in industry as well. For comparability and assessing the impact of occlusions during the image orientation phase, we include as well a scenario without obstacles, occlusions and visibility constraints. Our goal is to help the benchmark progress because there is no existing dataset dedicated to this problem setting. In addition, this is the first attempt to evaluate the geometry behind obstacles reconstructed by NeRFs through a point cloud comparison in 3D space, addressing accuracy and completeness. The key challenge lies in recovering the underlying occlusion and completing the invisible parts of the object, thus investigating if NeRFs can compete against traditional MVS for scenarios where the latter falls short. Considering that NeRFs' output is a continuous volumetric representation where the density field represents the geometry and it solely depends on the position, we propose a simple yet effective strategy to extract NeRFs' geometry in the form of a point cloud. We voxelize the whole density field and apply a 3D density-gradient based Canny edge detection filter to better represent the object's geometric features.

Our contributions can be summarized as follows:

(1) We introduce a new dataset named STELLA of real-world scenarios tackling transparent and non-transparent obstacles, occlusions and visibility constraints.
(2) We evaluate the underlying geometry of NeRFs for reconstructing the occluded object parts behind obstacles by comparing point clouds in 3D space with respect to ground truth.
(3) We propose a new approach for NeRFs geometry extraction in the form of a point cloud by voxelizing the density field and applying a 3D density-gradient based Canny edge detection filter to better constrain the geometric features.

This contribution is organized as follows: an outline of the related work concerning image-based 3D reconstruction using MVS and NeRFs as well as obstacle datasets is presented in Section 2. Following that, in Section 3 the processing steps for the 3D reconstruction and extraction of the point clouds from NeRFs along with the evaluation

metrics are described. Section 4 highlights the data acquisition setup, followed by the qualitative and quantitative results in Section 5. The discussion is laid out in Section 6 and Section 7 provides the final conclusions.

## 2. Related Work

In this section we give an overview of the respective literature. Namely, Section 2.1 refers to the traditional and learning-based MVS, then in Section 2.2 the neural implicit 3D geometric reconstruction methods are presented and Section 2.3 gives an overview of current research concerning occluded object parts and obstacle benchmark datasets.

### *2.1. Multi-View Stereo*

Traditional MVS methods rely on pixel matching algorithms to compute a 3D representation of the scene from multiple images. Based on the projection between a set of overlapping images taken from different positions, the depth values of each pixel are calculated, which is time-consuming and restricts the completeness of the reconstructed point cloud [9]. Since MVS often fails in estimating correct depth pixel values in low-textured areas of the scene, applying 3D edge extraction can support the mesh triangulation step [10] and preserve geometric details. However, the quality of the reconstructed mesh strongly depends on the accuracy of the extracted 3D edges.

Instead of relying on explicit geometric computations, learning-based MVS methods use neural networks for regularization and feature extraction to achieve superior performance against traditional methods [11]. One significant advantage is the ability to capture high-level semantics and contextual information, allowing for improved depth estimation on texture-less surfaces [12]. As a geometry-aware model [13], besides recovering per-pixel depth values [14], the neural network can learn to estimate the surface normal and reflectance of homogeneous and complex surfaces. By learning to infer the underlying scene structure based on available information from visible regions, these methods are also proficient in handling occlusions, noise, blur and distortion [15] through distinctive key points identification. However, for real-time applications the balance between accuracy and efficiency is crucial; hence, it sets limitations to the method's applicability.

### *2.2. Neural Implicit Scene Reconstruction*

Differing from traditional MVS, which reconstructs a scene by identifying common features using triangulation techniques to determine their geometric position, NeRFs represent scenes as a continuous volumetric field consisting of position-dependent density and view-dependent radiance. However, NeRFs assume controlled conditions and static bounded scenes and can be computationally demanding, particularly when processing high-resolution images. To address these drawbacks, researchers have been actively contributing to enhance the method's coherence, applicability and performance. Consequently, an extension of Mip-Nerf [16] that resolves the anti-aliasing issue and allows scene reconstruction on different scales, Mip-NeRF 360 [17] overcomes the challenges presented by unbounded real-world scenes with unconstrained camera orientations. NeRF-W [18], on the other hand, extends NeRFs' capabilities for creating 3D reconstructions in uncontrolled real-world environments. Through an effective alignment-aware training strategy, AligNeRF [19] improves NeRFs' performance on high-resolution data. A structure-aware 3D scene representation more efficient than NeRFs is introduced in Nerflets [20], which decompose a scene into multiple neural fields to better describe an object's appearance, density and semantics. Neuralangelo [21] utilizes multi-resolution 3D hash grids and neural surface rendering to achieve superior results in recovering dense 3D geometry from multi-view images, enabling highly detailed scene reconstruction. NeRF-RPN [22] develops a framework for 3D object extraction anisotropically, without class labels, by using only density values. Point-NeRF [23] and Points2NeRF [24] model a volumetric radiance field by combining the advantages from both traditional point cloud reconstruction and neural radiance field representation. Recent improvements focusing on accelerating the inference

step enable scene reconstruction within a few seconds, even with high-resolution images. More precisely, Instant Neural Graphic Primitives (Instant-NGP) [25] address the issue of computational efficiency while preserving rendering quality using a small neural network augmented by multi-resolution hash encoding, implemented on fully fused CUDA kernels.

Considering reflective surfaces, NeRFs produce inaccurate depth estimation due to the mixed geometry of the transmitted and reflected components [26]. Recently, some state-of-the-art methods tackle the reconstruction of the 3D geometry of transparent objects and the ability of seeing through the glass [27]. By assuming that the poses and geometry of a glass container are known and estimated in a controlled setup, the object's geometry inside the container can be accurately extracted. Other methods [28,29] focus on solid transparent objects and the geometry and refraction of light are treated separately. However, they require preprocessing for environmental light since the images are captured under static natural illumination in a controlled environment for one object per scene.

### *2.3. Occlusion Benchmark Datasets*

Recovering the geometry of objects behind obstacles is tackled by decomposing an image to background and front objects by using masks based on assumptions about the occluded parts [30]. However, this approach is only suitable if small parts of the scene are occluded and the objects have simple geometry. In addition, several benchmark datasets [31–36] focus on occluded parts behind obstacles in real-world scenes from non-overlapping images using single image annotations and masking for semantic extraction. Other researchers consider texture-less scenes with no reflective properties [37,38] captured in controlled lightning and a restricted number of images [39–41] to identify occlusions through object detection.

We can conclude that although the numerous variants of NeRFs have offered new insights in 3D reconstruction from images for real-world scenes, most are primarily focused on the task of novel-view synthesis; hence, the evaluation refers to the radiometric image quality. Moreover, a critical aspect still remains unexplored; evaluating the geometry of the object's occluded parts behind obstacles as there is no dataset depicting this type of environment. Current benchmark datasets tackling obstacles are not suited for image-based 3D reconstruction since they consist of non-overlapping images and thus are only suitable for semantic segmentation and object detection. In contrast, we bring the evaluation from image to 3D metric space through a point cloud comparison against traditional MVS, addressing the accuracy and completeness to investigate if NeRFs can challenge conventional dense matching algorithms in obstacle scenarios. Furthermore, our STELLA dataset concerns transparent and non-transparent obstacles with overlapping images, which enables a reconstruction and evaluation of the geometry in 3D space. We perform the evaluation against ground truth data, allowing for the development and comparison of the performance of different algorithms using the provided evaluation metrics.

## 3. Methodology

As depicted in Figure 1, in Section 3.1 the principles of pose estimation are introduced. Subsequently, Section 3.2 briefly summarizes the MVS dense matching principles while, in Section 3.3 the 3D reconstruction and extraction of the geometry as a point cloud from NeRFs are described. Lastly, Section 3.4 underlines the evaluation metrics.

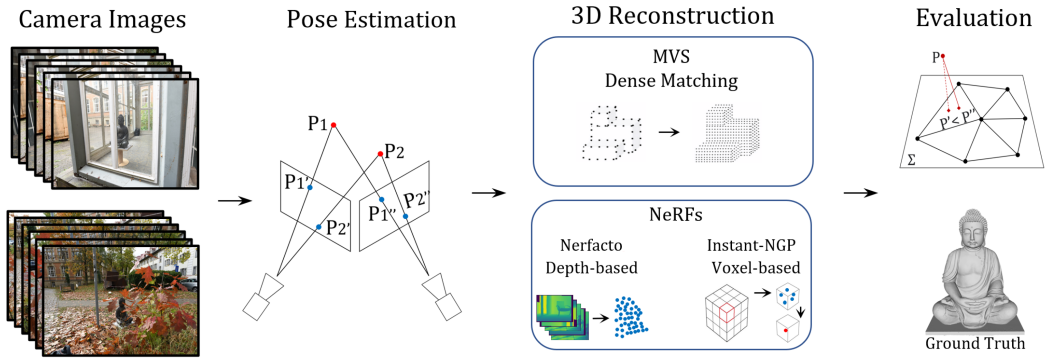



**Figure 1.** Flowchart of the undertaken investigations. The captured images with estimated camera poses (Section 3.1) are input for MVS (Section 3.2) and NeRFs (Section 3.3). Besides dense matching in MVS, we apply two point cloud extraction approaches for NeRFs, namely depth-based and voxel-based, to analyze how the point cloud generation and extraction approach affects the 3D geometric reconstruction. All point clouds are registered in the same metric space as the ground truth mesh for evaluation (Section 3.4).

### *3.1. Pose Estimation*

Camera poses are estimated independently through SfM. Namely, the algorithm expects as input a set of overlapping images captured from different camera positions. The camera poses along with the 3D geometry are determined by detecting robust distinctive key points in all images. Then, descriptors that capture the appearance and geometric properties of those points are generated, enabling robust image matching. The key points are matched over all images based on similarity, depicting the same 3D point. An object is recognized in a new image by individually comparing each feature from the new image with the extracted ones. These corresponding key points are input to a bundle adjustment that determines the camera poses and intrinsic parameters, as well as the 3D coordinates for all key points depicted in multiple images, resulting in a sparse point cloud.

### *3.2. Multi-View Stereo*

In order to investigate if NeRFs can compete against conventional dense matching algorithms in obstacle settings, a traditional MVS dense point cloud is used for comparison purposes. The core concept of MVS involves triangulation of corresponding features that are matched and tracked across multiple images to calculate the position of 3D points, hence creating a dense point cloud that captures the spatial structure of the scene. As MVS is used to refine the output obtained by SfM, it takes the information from the sparse point cloud for the pixel-wise computation of depth values per image. Since it operates in a more constrained environment with known camera parameters, it computes the 3D vertices on regions that could not be correctly detected by descriptors or matched in the previous step.

### *3.3. Neural Radiance Fields*

Reconstructing NeRFs' geometry in 3D space is achieved by predicting the depth given a series of images with estimated camera poses that describe different perspectives of an object. Nonetheless, the output is a continuous volumetric field and the geometry in the form of a point cloud needs to be accessed. Therefore, we use two different point cloud extraction approaches, namely depth-based and voxel-based, and analyze the results. Our focus is on evaluating how the point cloud extraction approach affects geometric accuracy and completeness.

**Nerfacto** We use Nerfacto in Nerfstudio [42] since it is mainly developed for real-world data captured outdoors for unbounded scenes. The method uses a small neural network with hash encoding for computational efficiency, while achieving comparable accuracy. Maintaining a dense point set for nearby objects is achieved by consolidating the points per ray into regions of the scene, usually with the first surface intersection. The

point cloud is extracted by rendering depth maps for each input image, back-projecting points from the depth maps and mapping to 3D coordinates with respect to the camera poses. The depth is derived from the expected ray termination in the density field.

**Instant-NGP** We use the original Instant-NGP framework since it achieves a similar accuracy as Nerfacto [43] but enables faster reconstruction. While the radiance field depends on viewing direction and does not separate color and illumination, the density field represents NeRFs geometry and is only related to query positions [44]. Consequently, we have integrated a .ply writer in order to extract the point cloud in voxel space by voxelizing solely the density field. Nevertheless, filtering with global density thresholds yields noisy and incomplete reconstruction [45,46]; thus, we apply a 3D density-gradient based Canny edge detection filter as it leads to higher accuracy and completeness [47]. Similarly, like in images where variations in magnitude depict edges, the aim of 3D edge detection is to locate edges belonging to boundaries of objects [48] by additionally taking into account the third dimension.

The first-order derivative is used to find out the minimum and maximum values in the gradient-based operator [49]:

$$\nabla f = G[f(x,y,z)] = \begin{bmatrix} Gx \\ Gy \\ Gz \end{bmatrix} = \begin{pmatrix} \frac{df}{dx} \\ \frac{df}{dy} \\ \frac{df}{dz} \end{pmatrix} \tag{1}$$

Then, the magnitude representing edge strength is calculated:

$$M(x,y,z) = |G| = \sqrt{(Gx)^2 + (Gy)^2 + (Gz)^2} = \sqrt{\left(\frac{df}{dx}\right)^2 + \left(\frac{df}{dy}\right)^2 + \left(\frac{df}{dz}\right)^2} \tag{2}$$

The gradient direction (orientation of edge normal) is given by:

$$\theta = \arctan\left(\frac{df}{dx}, \frac{df}{dy}, \frac{df}{dz}\right) \tag{3}$$

The gradient calculation in the voxelized 3D density field based on the density values is first preceded by Gaussian smoothing to minimize noise, followed by gradient magnitude thresholds on the density gradients and a hysteresis method to suppress weak and detect strong edges. The density gradients are derived independently of the absolute magnitude of density values, which allows for edge extraction along lower density values in the density field.

In spite of that, NeRF reconstructions have points inside the object since points are sampled along camera rays to render the volume density field. We strive to remove these points by approximating the visibility of the point cloud from a given viewpoint. Due to the object's complexity, we render five different lines of sight with regard to a given radius.

### *3.4. Evaluation Metrics*

Geometric evaluation is a critical aspect of assessing the quality of any 3D geometric reconstruction. The goal is to quantify how well the geometric attributes match the real-world features that they represent. For that reason, all point clouds are aligned in the same metric space as the ground truth mesh for evaluation using the Iterative Closest Point (ICP) [50], which finds an optimal rigid transformation to align two point sets. To evaluate the geometry of the reconstructed point clouds, we use two criteria: accuracy (precision) and completeness (recall).

**Accuracy** Accuracy quantifies how closely the reconstructed point cloud reflects ground truth locations. We use cloud-to-mesh, which computes the displacements between each point in the compared point cloud and the nearest facet in the reference mesh through Euclidean distance. The orthogonal (signed) distance from the point to the nearest triangle plane, or the distance to the nearest edge is taken if the orthogonal projection of

the point does not fall on any facet. Due to the normal distribution without long-tailed data, measuring how dispersed the data is in relation to the ground truth mesh is provided through Mean Error (Mean), Standard Deviation (SD) and Root Mean Square Error (RMSE) [51] accordingly:

$$Mean = \frac{\sum_{i=1}^{n}(x_i)}{n} \tag{4}$$

$$SD = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}} \tag{5}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}} \tag{6}$$

where $n$ is the number of points of the compared point cloud, $x_i$ stands for the closest distance between each point in the compared point cloud and the nearest facet of the reference mesh, while $\overline{x}$ denotes the mean value of the distances.

**Completeness** Completeness measures to what extent the ground truth surface is covered. We report both qualitative and quantitative completeness by calculating the percentages of points covered by MVS and NeRF reconstructions in all scenarios. A higher percentage indicates higher completeness.

## 4. Datasets

We base our investigations on a real-world dataset STELLA, consisting of three scenarios: without obstacles, occlusions and visibility constraints (*Original*) and with two types of obstacles, transparent (*Window*) and non-transparent (*Vegetation*). The object behind the obstacles whose recovered geometry should be evaluated is a 0.7 m tall Buddha statue (further on referred to as object) placed on a 0.48 × 0.38 × 0.02 m rectangular plate (Figures 2 and 3).

(**a**) (**b**)

**Figure 2.** Image capturing setup for both obstacle scenarios with Nikon D810 SLR mounted on a tripod: (**a**) We place the object behind windows whose glass isn't completely transparent, it has dust and paint dots which make the reconstruction challenging; (**b**) Different vegetation coverage is considered in the front side of the object.

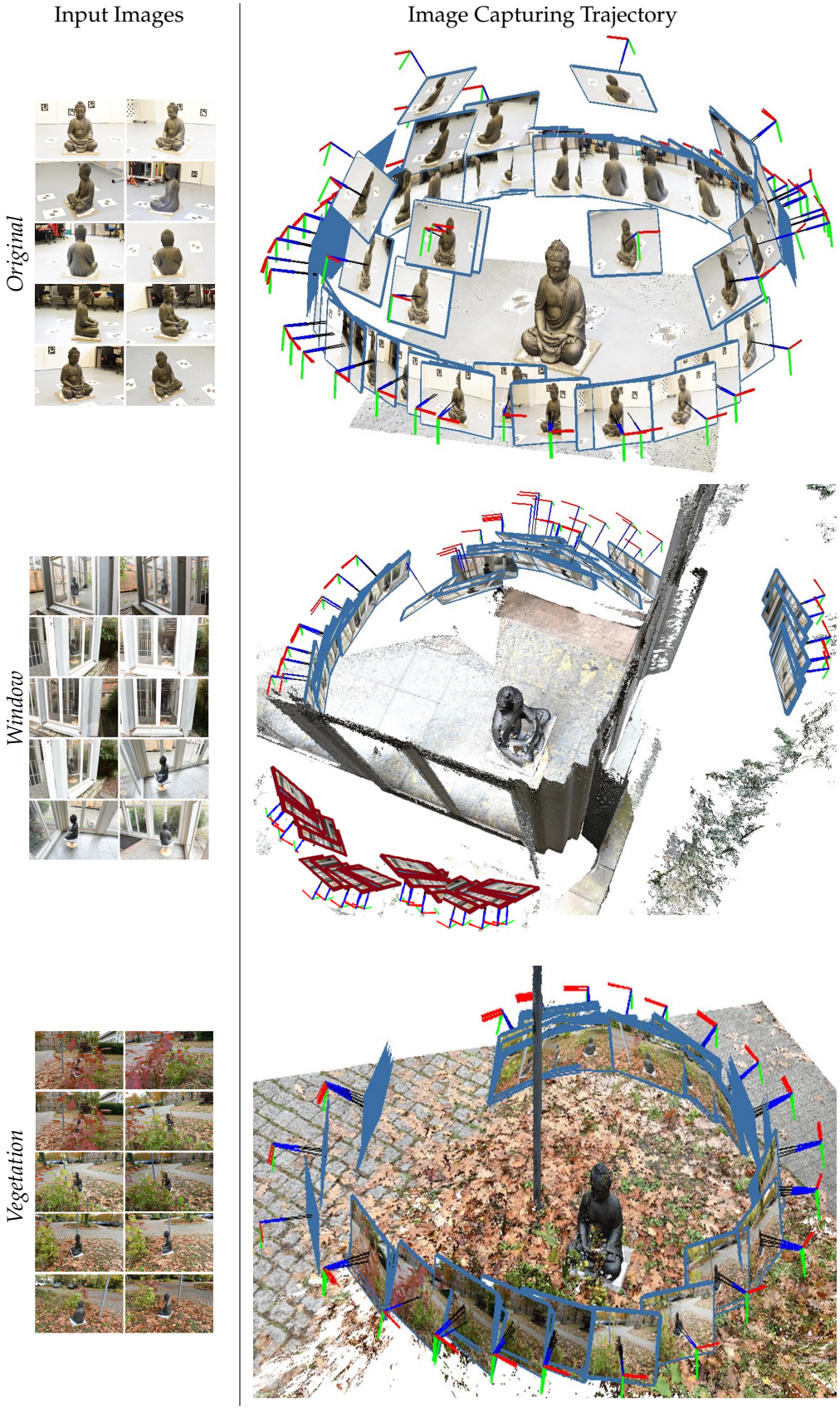
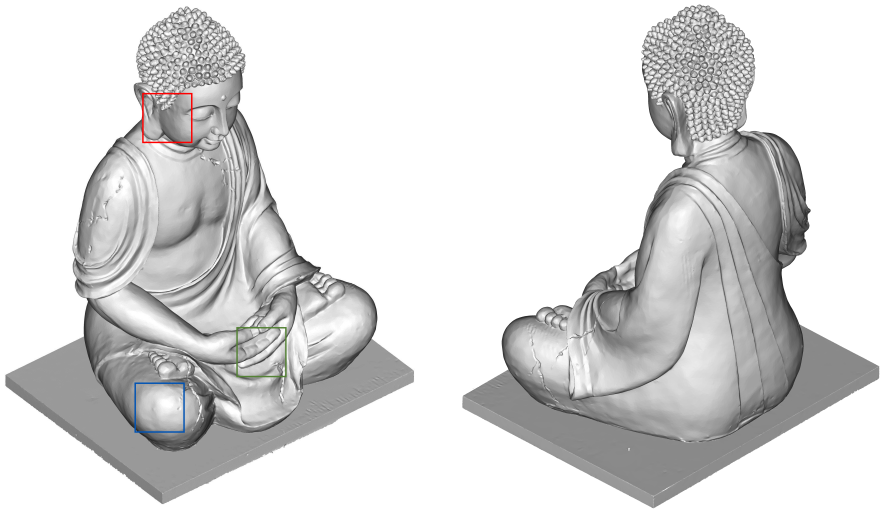



**Figure 3.** Trajectory of the image capturing with visualized camera poses for all scenarios: *Original*, *Window* and *Vegetation*. The red squares indicate that the images are not taken for the pose estimation due to homogeneous background and repetitive texture.

**Image capturing** Image-based 3D reconstruction is sensitive to variations in image quality as images with low resolution, motion blur or noise can hinder the accuracy of feature matching and depth estimation. Therefore, all scenarios are captured using a Nikon D810 SLR digital camera with a 36MP sensor, image resolution of 7360 × 4912 pixels, 20 mm focal length and f/8 aperture size. Due to the high resolution, the camera is mounted on a tripod (Figure 2) to prevent shivering and vibrations during acquisition. The images for the *Original* scenario without obstacles, occlusions and visibility constraints are captured indoors in a hemispheric trajectory around the object from three different camera heights of 0.7 m, 1.1 m and 1.5 m respectively, to achieve full object coverage. For the *Window* scenario, the object is placed behind two windows (front and right side) whose glass is not completely transparent, it has dust and paint dots to additionally challenge the pose estimation, hence the reconstruction. For the *Vegetation*, different types of vegetation coverage are considered in the front side of the object. The images are captured in a circular trajectory from a uniform distance of the object (Figure 3). Each scenario consists of 125 images which are used in our research purposes.

**Ground truth** The ground truth data in the form of a mesh are obtained using Structured Light Imaging (SLI). The method captures the geometry of the object by illuminating the surface using structured light projection. 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 can then be triangulated from the intrinsic and extrinsic camera parameters and the image coordinates. For this purpose, we utilize the stereoSCAN scanning device whose projector is placed between two digital cameras at a fixed distance from each other. The object is placed on a turntable and the rotations are controlled automatically by a workstation [52]. Any remaining holes are closed, resulting in a smoothed mesh with 0.1 mm accuracy (Figure 4).



(**a**) (**b**)

**Figure 4.** Structured Light Imaging (SLI) 3D mesh as ground truth: (**a**) front and (**b**) back view. The colored squares depicting the geometry of the occluded object parts are enlarged in Figure 6.

## 5. Results

In this section, we conduct evaluations on our STELLA dataset, which contains three scenarios: without line-of-sight obstructions (*Original*) and with transparent (*Window*) and non-transparent (*Vegetation*) obstacles against a ground truth mesh. To evaluate how the point cloud generation and extraction approach affects the 3D geometric reconstruction, we compare the dense matching by MVS with two NeRF approaches, namely depth-based and voxel-based, and argue which method represents occlusions more accurately and reliably. In order to quantify the quality of the reconstructed geometry, we calculate accuracy and

completeness between the point clouds and the ground truth mesh. For accuracy, we report three error metrics: Mean, SD and RMSE, while the completeness is presented as a percentage. The inputs to all methods are 125 high-resolution images for each scenario. The results are categorized as qualitative (Section 5.1), where the geometric reconstructions as point clouds are visualized (Figures 5–7), and quantitative (Section 5.2), to numerically present the accuracy and completeness across all scenarios respectively (Table 1).


**Figure 5.** Qualitative comparison of the point cloud geometric reconstructions for all used methods: MVS, Nerfacto and Instant-NGP in all scenarios: without obstacles, occlusions and visibility constraints (*Original*), transparent (*Window*) and non-transparent (*Vegetation*) obstacles respectively. MVS shows gaps in reconstructing the geometry of the occluded parts, while NeRFs exhibit in completeness.

### *5.1. Qualitative Results*

The qualitative comparison of the point clouds for the used methods in all scenarios in Figure 5 highlights the reconstruction quality of the geometry and object completeness. Since we are investigating the 3D reconstruction through obstacles, we focus on the geometry of the occluded parts of the object.

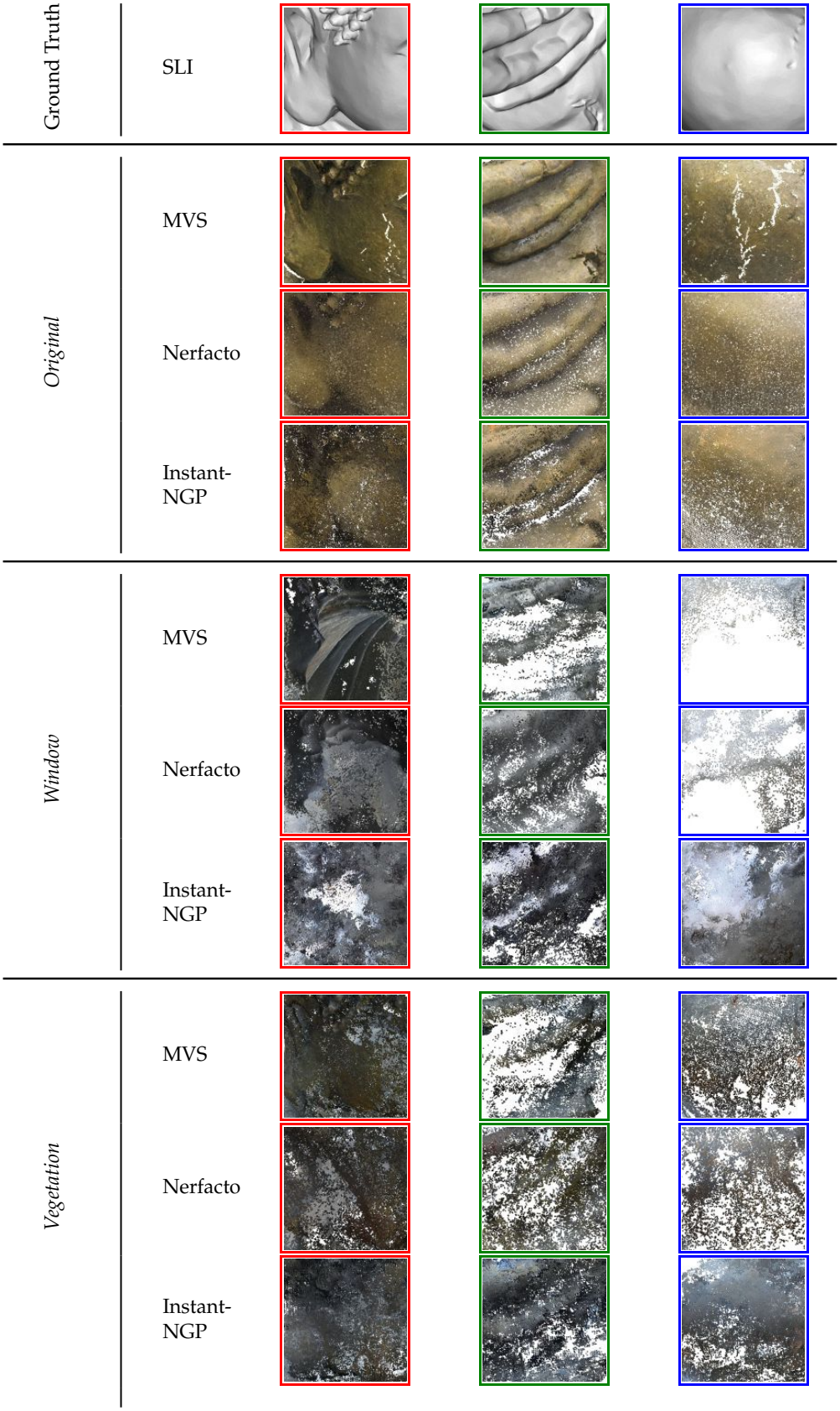



**Figure 6.** Enlargements of the occluded parts for all used methods in all scenarios against the ground truth (Figure 4). MVS manifests significant gaps in the reconstruction, while NeRFs achieve higher object completeness and better reconstruct the underlying geometry behind obstacles.

**Accuracy** The error displacement values against the ground truth mesh correspond to the color ramp, where green indicates errors close to 0 (Figure 7). First, the approximate distances are computed, which are used to set the best octree level at which the distance computation is performed. All distances above the mesh have positive values, while negative distances are an indication that the point lies below the nearest triangle. Following the suggestion of the initial NeRF paper and because other commercial alternatives do not lead to better geometric representation, MVS is reconstructed using COLMAP [3], and even though it fails to reliably reconstruct the object's surface behind the occluded areas, when it comes to accuracy, it shows almost a uniform match with the ground truth. In contrast, NeRF reconstructions in all scenarios show a fuzzier interpretation of the object's outer surface because there are not sufficient surface constraints in the implicit representation. The density is learned implicitly and is not conditioned by ground truth information since the network optimization is based on minimizing the image reconstruction loss between training and rendered images [53]. Moreover, NeRFs' accuracy is strongly dependent on the point cloud extraction approach. Nerfacto shows higher correspondence with the ground truth since the point cloud is extracted directly from the depth maps. On the contrary, Instant-NGP point clouds contain artifact points within the object (*Vegetation*) because NeRFs sample points in the entire 3D space and we are addressing the whole voxelized density field. Those points have the highest error displacements and significantly distort the accuracy results.

**Completeness** Generally, MVS achieves a sharper result and complex geometry like the spikes on the head and the facial features are reliably reconstructed (Figure 5). However, the gaps in the geometry of the occluded parts decrease the completeness. On the other hand, we remark that NeRFs can capture the overall spatial arrangement of the scene completely while being able to also capture geometric details, e.g., the spikes on the head and other surface details (Figure 6) of occluded object parts, thus achieving higher completeness. For the *Original* and *Vegetation* scenario, Instant-NGP and our voxel-based point cloud extraction approach provide the strongest visual results. However, noise and artifacts can be observed in the reconstructions, particularly in Instant-NGP reconstruction for the *Window* scenario. NeRF artifacts arise because the method cannot capture the object's accurate and clean geometry resulting in floaters, and cannot complete the neural density field resulting in holes. NeRFs demand hundreds of input images with highly accurate camera poses and fall short under conditions of sparse observations, especially in real-world scenarios, resulting in a foggy density field. The camera poses are an essential part of NeRFs because points are sampled per ray emitted from the camera's center along the pixel's direction. Misaligned camera poses result in cloudy artifacts and a reduction in the sharpness and details of the reconstructed scene. Even with high-resolution input images, the geometry varies based on the scenario's constraints, such as acquisition constellations, types of obstacles and occlusions. Hence, the completeness as well as the accuracy are highest in the *Original* scenario because the scene is bounded, background content does not exist in any direction and the images are captured on a surrounding hemisphere (Figure 3), covering a wide range of object's surface.

We can also observe discoloration among different scenarios, namely the visual appearance of the *Original* scenario strongly differs from the *Window* and *Vegetation*. 3D geometric reconstruction is sensitive to lighting conditions and materials. Variations in lighting can lead to inconsistent color and texture information in images. The *Original* scenario is captured indoors under controlled lighting, which is different from the other two obstacle scenarios, captured outdoors under natural illumination, which is not constant. In NeRFs, the color of a point does not change when viewed from different directions; instead, it gets determined as a weighted sum of all spatial points along the ray in the radiance field independently of the volume density. However, apart from visual appearance, the color differences do not affect the geometric reconstruction.

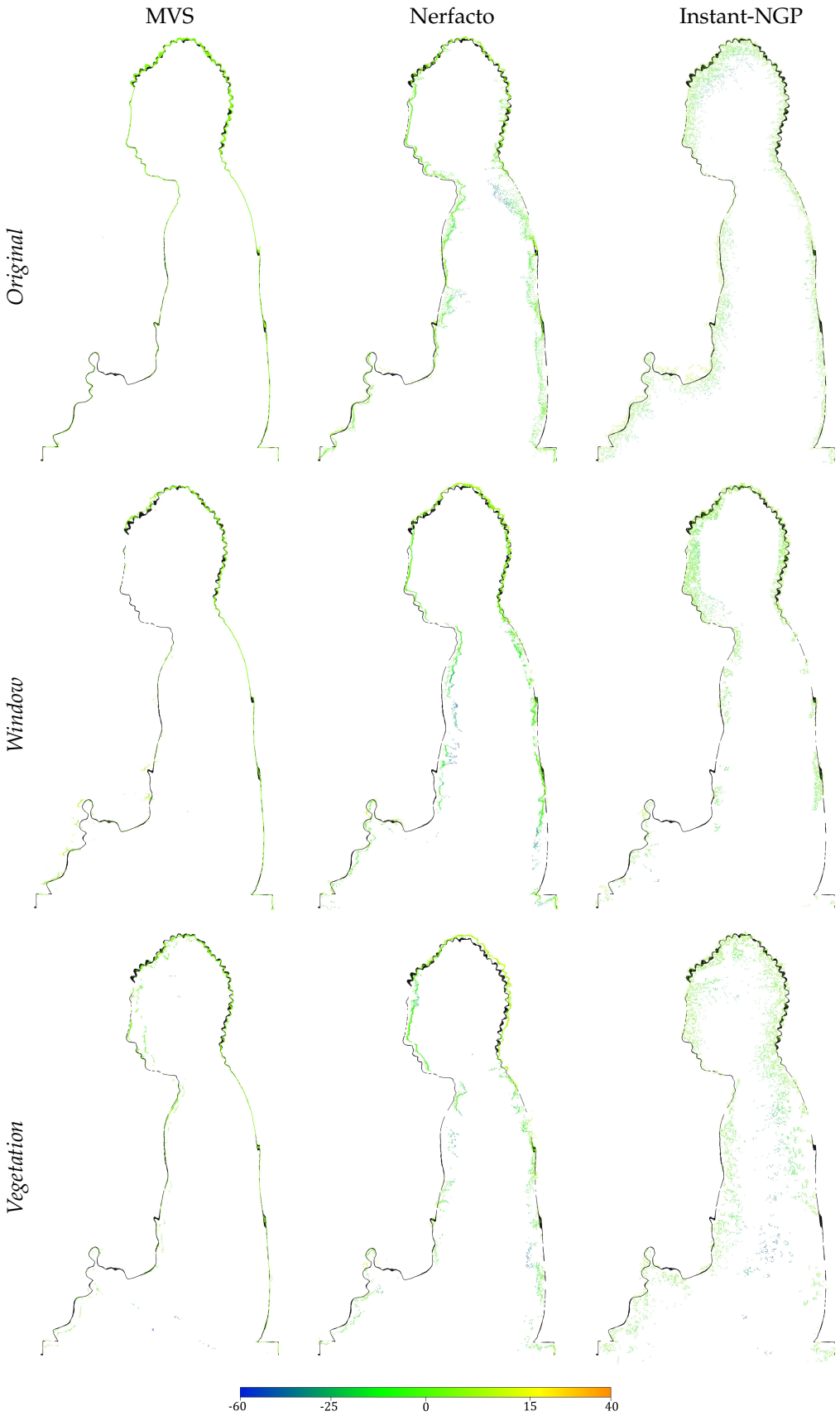



**Figure 7.** Cloud-to-mesh comparison by visualization of the reconstructed point clouds against the ground truth accordingly for all scenarios, *Original*, *Window* and *Vegetation*. The ground truth mesh is presented as a thin black line and the colors correspond to the error displacements (in mm) in a longitudinal section of the object.

### 5.2. Quantitative Results

Table 1 provides a quantitative comparison of the used methods for 3D reconstruction: MVS, Nerfacto and Instant-NGP in all scenarios: *Original*, *Window* and *Vegetation* and underscores the trade-off between accuracy and completeness among the methods.

**Accuracy** MVS demonstrates superior accuracy among the methods across all scenarios with lowest error displacements in all metrics, thus achieving the highest correspondence with the ground truth. In the *Original* scenario, where the line of sight is unobstructed, the average difference between the reconstructed point cloud and the ground truth mesh is a remarkable 0.41 mm. Even though the divergence values for both *Window* and *Vegetation* obstacle scenarios are up to 2.27 mm and 3.49 mm respectively, MVS indicates almost a uniform match with the real-world features it represents. It should be noted that although NeRFs provide lower geometric accuracy compared to MVS, the method shows more robust results as the accuracy values are within the same range for each scenario; however, they are strongly dependent on the chosen point cloud extraction approach. More precisely, Nerfacto using depth rendering to extract the explicit geometry, achieves accuracy within 8.52 mm and 8.49 mm, respectively, in both obstacle scenarios for most parts of the object surface. Instant-NGP shows higher divergence from the ground truth, with 18.66 mm for the *Window* scenario and 16.82 mm for the *Vegetation* scenario, since points are sampled in the entire volumetric representation and for the point cloud extraction we have voxelized the whole density field.

**Completeness** Although MVS outperforms in pinpoint accuracy, it fails to reconstruct the object's outer surface, especially the occluded geometry behind obstacles on account of completeness. In all scenarios, the reported values are lower compared to NeRFs. The absolute completeness values range between approximately 65 and 77% for both obstacle scenarios respectively, proportionally with the number of points. Regardless of the point cloud extraction approach, NeRFs consistently provide a higher point coverage with respect to the ground truth across all scenarios, suggesting a more robust reconstruction through obstacles in diverse environments. However, between NeRF reconstructions, our Instant-NGP voxel-based point cloud extraction approach outperforms the depth-based method by Nerfacto, with completeness scores within 94% and 84% for the *Original* and *Vegetation* scenario respectively.

**Table 1.** Quantitative results of the cloud-to-mesh comparison addressing the number of points, accuracy and completeness for each method. MVS outperforms both NeRFs point cloud extraction approaches when it comes to accuracy. However, NeRFs show higher coverage for the ground truth points. The best results are highlighted in **bold**.

| Scenario | Method | Number of Points (Million) | Accuracy (mm) | | | Completeness (%) ↑ |
|---|---|---|---|---|---|---|
| | | | Mean ↓ | SD ↓ | RMSE ↓ | |
| *Original* | MVS | 16.48 | **0.02** | **0.40** | **0.41** | 88.96 |
| | Nerfacto | 2.65 | 3.78 | 5.29 | 6.50 | 90.12 |
| | Instant-NGP | 2.53 | 3.51 | 6.61 | 7.48 | **93.65** |
| *Window* | MVS | 2.66 | **0.19** | **2.26** | **2.27** | 65.26 |
| | Nerfacto | 3.11 | −3.9 | 7.58 | 8.52 | **83.38** |
| | Instant-NGP | 2.98 | −10.10 | 15.69 | 18.66 | 80.92 |
| *Vegetation* | MVS | 1.92 | **−0.95** | **3.36** | **3.49** | 76.94 |
| | Nerfacto | 1.72 | −3.43 | 7.76 | 8.49 | 69.59 |
| | Instant-NGP | 2.16 | −7.82 | 14.89 | 16.82 | **83.99** |

## 6. Discussion

In this contribution we evaluate the 3D geometry reconstructed by NeRFs against traditional MVS, tackling the accuracy and completeness of the occluded object surface behind obstacles. We can notice a clear trade-off between accuracy and completeness among the methods; while MVS consistently exhibits higher accuracy, NeRFs outperform in completeness, indicating a consistent performance across different obstacle scenarios.

### *6.1. Qualitative Evaluation*

**Accuracy** MVS achieves the best geometric accuracy results with highest correspondence with the ground truth (Table 1) and the 3D geometry is reconstructed by identifying common image features, using triangulation techniques to determine their 3D position. NeRFs, on the other hand, use deep learning to infer a continuous volumetric representation from a set of overlapping images without identification of common features, but rather learn an implicit function that maps 3D coordinates along the rays. The volume density is essentially inferred as part of this process since it affects how much light is absorbed and scattered along the ray between the 3D point and the camera viewpoint. However, NeRFs are trained by minimizing the image reconstruction loss over training views through gradient descent. The volume density field is thus flexible so each training can lead to a slightly different geometry prediction and result. The error displacements tend to increase faster to the negative because those are the points behind the mesh surface, caused by the fact that NeRFs sample points in the whole 3D field and consequently inside the object. The errors with positive values lie above the nearest mesh triangle and thus most likely represent noise and outlier points.

**Completeness** Reasonably, the incomplete geometry in MVS and NeRFs occurs in the occluded areas behind the obstacles (Figure 5). MVS shows bigger gaps in reconstructing the object's surface and less complete geometry (Figure 6) but a sharper result. On the contrary, NeRFs generally show a more reliable reconstruction of the occluded parts and demonstrate higher points coverage. Generally, the empty spaces in a point cloud stem from two main sources: an insufficient number of overlapping images and constraints of the reconstructive algorithm. When the camera poses are known, the dense point cloud is generated using the location and orientation information from SfM. The quality of image-based 3D reconstruction is highly dependent on the presence of distinctive features and texture in the images to establish correspondences between points for the pose estimation. Scenes with uniform or repetitive patterns can enforce challenges as feature extraction and matching may be less effective to accurately reconstruct the 3D geometry. On top of that, light paths through unpolarized glass are complex and involve reflection. These multiple reflections cause image distortions that invalidate the single viewpoint assumption as a different reflection might be shown depending on the viewpoint. Hence, it imposes difficulties in reconstructing the 3D geometry of such objects because the presence of specular and mirror reflections can lead to noisy results. The glass in our *Window* scenario is not completely transparent, it has filth and paint dots which additionally challenge the pose estimation and subsequently the reconstruction. As a consequence, the images (30 in total) through the glass from the right side of the object are not taken for pose estimation (Figure 3), resulting in gaps (right arm and face), especially for MVS since it strongly depends on the object's visibility on numerous overlapping images for a reliable reconstruction. In contrast, NeRFs show more complete geometry in those parts as the geometry is reconstructed through ray tracing from the camera poses and 3D point sampling, particularly Instant-NGP, which almost completely approximates the missing parts. However, the point cloud is quite noisy and the facial features are poorly reconstructed. NeRFs can often produce artifacts due to the inherent ambiguity of the task of reconstructing an unbounded scene from a relatively small set of images.

Without affecting the object's geometry, we can identify color differences in the reconstructed point clouds (Figure 5), knowing that the color of the object may change depending on the viewpoint. The radiance field is sampled along camera rays and the samples are

aggregated to predict the pixel color using hierarchical stratified sampling, meaning that the ray is divided into equally spaced bins, where a sample is uniformly drawn from each bin. Using stratified samples generally improves the reconstructions as it helps to prevent overfitting. Once the space is sampled, it is known which samples contributed to the final color, which can be used to sample more points around those regions, leveraging a probabilistic function to determine the expected color value along the ray. NeRFs do not explain the view dependency of a reflecting surface point by changing the color of a point when viewed from different directions, but they utilize all spatial points behind it along the camera ray to obtain the correct color, resulting in incorrect depth maps [26]. The visual appearance of the *Original* scenario strongly differs from the *Window* and *Vegetation* scenarios. The *Original* scenario is captured indoors under different light compared to the other two obstacle scenarios, captured outdoors under natural illumination, which can vary. Moreover, due to the transparency in the *Window* scenario, both NeRF point clouds show minor discoloration caused by the reflecting properties of the glass. However, mixed color points are more present in the *Vegetation* scenario, especially for Nerfacto, considering the non-transparency of the vegetation as an obstacle. Instant-NGP shows better visual results since we voxelize only the density field that does not consider the color as it is only position-dependent.

### *6.2. Quantitative Evaluation*

**Accuracy** NeRFs' accuracy strongly depends on the point cloud extraction approach. This is especially evident for Instant-NGP reconstructions which contain artifact points inside the object (Figure 7) that decrease the accuracy, because we have voxelized the whole density field. Unlike MVS, whose accuracy drops from 2.27 mm for the transparent (*Window*) obstacle to 3.49 mm for the non-transparent (*Vegetation*) obstacle, NeRFs' accuracy is almost constant in both obstacle scenarios, indicating a more robust reconstruction through obstacles in diverse environments. Additionally, even though we have rendered five different lines of sight due to the object's complexity, when applying the point removal, a smaller radius has been set because we do not want to sacrifice the completeness. The point removal approach is viewpoint-dependent and better suited for symmetrical geometric forms. It is worth mentioning that the point distribution on the object's surface is not proportional because the spikes on the head and the facial features contain more points. It is likely that those points have higher density values since NeRFs are optimized to capture complex geometry with high level of detail.

The *Original* scenario achieves the best accuracy and completeness results for all methods compared with the two obstacle scenarios owing to the more accurate pose estimation and object visibility on all images. The camera parameters are taken to compute the depth information for every pixel in an image. Fusion of the depth maps of multiple images in 3D then produces a dense point cloud to recover the 3D geometry of the scene. Thus, incorrect camera poses lead to incorrect depth maps and strongly affect the geometric accuracy and subsequently the completeness. Assessing the impact of occlusions during the image orientation requires a more synthetic environment while keeping the same scene setup and adding different obstacles. However, in this contribution we are focused on real-world challenging settings to address industry audiences as well. Obtaining an accurate geometric representation from NeRFs still remains an open challenge since NeRFs are optimized for view synthesis and do not enforce an accurate underlying geometry on the density field.

**Completeness** The completeness percentage is proportional to the number of points, except in the *Original* scenario, where MVS in spite of the fact that it has 16.48 million points, still exhibits the lowest completeness because the points are not proportionally distributed on the object's surface. For comparability, the point clouds should have a similar number of points; however, this parameter cannot be adjusted in COLMAP. It is calculated through cross-view correspondence matching and triangulation to estimate the pixel-wise depth values and the input images have a high resolution of $7360 \times 4912$ pixels.

## 7. Conclusions

In this contribution we evaluate the 3D geometry of NeRFs against traditional MVS for reconstructing objects through transparent and non-transparent obstacles against a ground truth. In addition, we introduce a new "obSTaclE, occLusion and visibiLity constrAints" STELLA (available online at https://github.com/sqirrel3/STELLA, accessed on 12 January 2024) dataset consisting of real-world challenging scenarios tackling transparent and non-transparent obstacles because there is no existing dataset dedicated to this problem setting. For comparability and assessing the impact of occlusions on the pose estimation, we include as well a scenario without visibility obstructions. Considering that the density field represents NeRFs' geometry and it solely depends on the position, we propose an effective strategy to extract NeRFs' geometry in the form of a point cloud by voxelizing the whole density field and applying 3D density-gradient based Canny edge detection to better represent the object's geometric features. We investigate how the point cloud generation and extraction approach affects the 3D geometric reconstruction and thus the accuracy and completeness. NeRFs' geometric accuracy is strongly dependent on the point cloud extraction approach and while it is lower compared to MVS, the method shows more robust results as the accuracy values are within the same range for each scenario. Nerfacto shows higher correspondence with the ground truth. On the contrary, Instant-NGP point clouds contain artifact points within the object that distort the accuracy results because NeRFs sample points in the entire 3D space and we are addressing the whole voxelized density field. However, NeRFs and our voxel-based point cloud extraction approach consistently provide a higher point coverage and hence outperform in completeness.

NeRFs still produce inconsistent geometry among multi-view observations, which is manifested into foggy floaters hovering within the volumetric representation. Obtaining an accurate geometric representation from NeRFs still remains an open challenge since the method is optimized for view synthesis and does not enforce an accurate underlying geometry on the density field. The density should be limited to only have positive values in the first intersection between the ray and object surface. Furthermore, NeRF reconstruction highly depends on accurately estimated camera poses, so investigating a different number of input images and degrading image quality can quantify the impact of the camera parameters on the geometric reconstruction.

**Author Contributions:** Conceptualization, I.P. and B.J.; Methodology, I.P. and B.J.; Software, I.P.; Validation, I.P.; Investigation, I.P.; Data curation, I.P.; Writing—original draft, I.P.; Visualization, I.P.; Supervision, B.J. All authors have read and agreed to the published version of the manuscript.

**Funding:** We acknowledge support by the KIT-Publication Fund of the Karlsruhe Institute of Technology.

**Data Availability Statement:** The dataset STELLA is available online at https://github.com/sqirrel3/STELLA (accessed on 12 January 2024).

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Dumic, E.; da Silva Cruz, L.A. Subjective Quality Assessment of V-PCC-Compressed Dynamic Point Clouds Degraded by Packet Losses. *Sensors* **2023**, *23*, 5623. [CrossRef] [PubMed]
2. Liu, Y.; Yang, Q.; Xu, Y.; Yang, L. Point cloud quality assessment: Dataset construction and learning-based no-reference metric. *ACM Trans. Multimed. Comput. Commun. Appl.* **2023**, *19*, 1–26. [CrossRef]
3. Schönberger, J.L.; Zheng, E.; Frahm, J.M.; Pollefeys, M. Pixelwise view selection for unstructured multi-view stereo. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Proceedings, Part III 14; Springer: Berlin/Heidelberg, Germany, 2016; pp. 501–518.
4. Schonberger, J.L.; Frahm, J.M. Structure-from-motion revisited. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 4104–4113.
5. Stathopoulou, E.K.; Battisti, R.; Cernea, D.; Remondino, F.; Georgopoulos, A. Semantically Derived Geometric Constraints for MVS Reconstruction of Textureless Areas. *Remote Sens.* **2021**, *13*, 1053. [CrossRef]

6. Sitzmann, V.; Zollhöfer, M.; Wetzstein, G. Scene representation networks: Continuous 3D-structure-aware neural scene representations. In Proceedings of the Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019.
7. Mildenhall, B.; Srinivasan, P.P.; Tancik, M.; Barron, J.T.; Ramamoorthi, R.; Ng, R. Nerf: Representing scenes as neural radiance fields for view synthesis. *Commun. ACM* **2021**, *65*, 99–106. [CrossRef]
8. Yan, Z.; Li, C.; Lee, G.H. Nerf-ds: Neural radiance fields for dynamic specular objects. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 8285–8295.
9. Zhang, X.; Yang, F.; Chang, M.; Qin, X. MG-MVSNet: Multiple granularities feature fusion network for multi-view stereo. *Neurocomputing* **2023**, *528*, 35–47. [CrossRef]
10. Stathopoulou, E.K.; Rigon, S.; Battisti, R.; Remondino, F. Enhancing Geometric Edge Details in MVS Reconstruction. *Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci.* **2021**, *43*, 391–398. [CrossRef]
11. Zhu, Q.; Min, C.; Wei, Z.; Chen, Y.; Wang, G. Deep learning for multi-view stereo via plane sweep: A survey. *arXiv* **2021**, arXiv:2106.15328.
12. Zhang, Y.; Zhu, J.; Lin, L. Multi-View Stereo Representation Revist: Region-Aware MVSNet. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023; pp. 17376–17385.
13. Zhang, Z.; Peng, R.; Hu, Y.; Wang, R. GeoMVSNet: Learning Multi-View Stereo With Geometry Perception. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023; pp. 21508–21518.
14. Yamashita, K.; Enyo, Y.; Nobuhara, S.; Nishino, K. nLMVS-Net: Deep Non-Lambertian Multi-View Stereo. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 2–7 January 2023; pp. 3037–3046.
15. Ito, K.; Ito, T.; Aoki, T. PM-MVS: PatchMatch multi-view stereo. *Mach. Vis. Appl.* **2023**, *34*, 32. [CrossRef]
16. Barron, J.T.; Mildenhall, B.; Tancik, M.; Hedman, P.; Martin-Brualla, R.; Srinivasan, P.P. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 5855–5864.
17. Barron, J.T.; Mildenhall, B.; Verbin, D.; Srinivasan, P.P.; Hedman, P. Mip-NeRF 360: Unbounded Anti-Aliased Neural Radiance Fields. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 5470–5479.
18. Martin-Brualla, R.; Radwan, N.; Sajjadi, M.S.; Barron, J.T.; Dosovitskiy, A.; Duckworth, D. Nerf in the wild: Neural radiance fields for unconstrained photo collections. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 7210–7219.
19. Jiang, Y.; Hedman, P.; Mildenhall, B.; Xu, D.; Barron, J.T.; Wang, Z.; Xue, T. AligNeRF: High-Fidelity Neural Radiance Fields via Alignment-Aware Training. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 46–55.
20. Zhang, X.; Kundu, A.; Funkhouser, T.; Guibas, L.; Su, H.; Genova, K. Nerflets: Local radiance fields for efficient structure-aware 3d scene representation from 2d supervision. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 8274–8284.
21. Li, Z.; Müller, T.; Evans, A.; Taylor, R.H.; Unberath, M.; Liu, M.Y.; Lin, C.H. Neuralangelo: High-Fidelity Neural Surface Reconstruction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 8456–8465.
22. Hu, B.; Huang, J.; Liu, Y.; Tai, Y.W.; Tang, C.K. NeRF-RPN: A general framework for object detection in NeRFs. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 23528–23538.
23. Xu, Q.; Xu, Z.; Philip, J.; Bi, S.; Shu, Z.; Sunkavalli, K.; Neumann, U. Point-nerf: Point-based neural radiance fields. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 5438–5448.
24. Zimny, D.; Trzciński, T.; Spurek, P. Points2nerf: Generating neural radiance fields from 3D point cloud. *arXiv* **2022**, arXiv:2206.01290.
25. Müller, T.; Evans, A.; Schied, C.; Keller, A. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Trans. Graph. (ToG)* **2022**, *41*, 1–15. [CrossRef]
26. Guo, Y.C.; Kang, D.; Bao, L.; He, Y.; Zhang, S.H. NeRFReN: Neural Radiance Fields with Reflections. *arXiv* **2022**, arXiv:cs.CV/2111.15234.
27. Tong, J.; Muthu, S.; Maken, F.A.; Nguyen, C.; Li, H. Seeing Through the Glass: Neural 3D Reconstruction of Object Inside a Transparent Container. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 12555–12564.
28. Wang, D.; Zhang, T.; Süsstrunk, S. NEMTO: Neural Environment Matting for Novel View and Relighting Synthesis of Transparent Objects. *arXiv* **2023**, arXiv:cs.CV/2303.11963.
29. Li, Z.; Long, X.; Wang, Y.; Cao, T.; Wang, W.; Luo, F.; Xiao, C. NeTO:Neural Reconstruction of Transparent Objects with Self-Occlusion Aware Refraction-Tracing. *arXiv* **2023**, arXiv:cs.CV/2303.11219.
30. Zhan, X.; Pan, X.; Dai, B.; Liu, Z.; Lin, D.; Loy, C.C. Self-supervised scene de-occlusion. In Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 3784–3792.

31. Zhou, Q.; Wang, S.; Wang, Y.; Huang, Z.; Wang, X. Human de-occlusion: Invisible perception and recovery for humans. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 3691–3701.
32. Zhang, S.; Xie, Y.; Wan, J.; Xia, H.; Li, S.Z.; Guo, G. Widerperson: A diverse dataset for dense pedestrian detection in the wild. *IEEE Trans. Multimed.* **2019**, *22*, 380–393. [CrossRef]
33. Zhuo, J.; Chen, Z.; Lai, J.; Wang, G. Occluded person re-identification. In Proceedings of the 2018 IEEE International Conference on Multimedia and Expo (ICME), San Diego, CA, USA, 23–27 July 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1–6.
34. Jia, M.; Cheng, X.; Lu, S.; Zhang, J. Learning disentangled representation implicitly via transformer for occluded person re-identification. *IEEE Trans. Multimed.* **2022**, *25*, 1294–1305. [CrossRef]
35. Ouyang, W.; Wang, X. A discriminative deep model for pedestrian detection with occlusion handling. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 3258–3265.
36. Lee, H.; Park, J. Instance-wise occlusion and depth orders in natural scenes. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–20 June 2022; pp. 21210–21221.
37. Hodan, T.; Haluza, P.; Obdržálek, Š.; Matas, J.; Lourakis, M.; Zabulis, X. T-LESS: An RGB-D dataset for 6D pose estimation of texture-less objects. In Proceedings of the 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), Santa Rosa, CA, USA, 24–31 March 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 880–888.
38. Tyree, S.; Tremblay, J.; To, T.; Cheng, J.; Mosier, T.; Smith, J.; Birchfield, S. 6-DoF pose estimation of household objects for robotic manipulation: An accessible dataset and benchmark. In Proceedings of the 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Kyoto, Japan, 23–27 October 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 13081–13088.
39. Blok, P.M.; van Henten, E.J.; van Evert, F.K.; Kootstra, G. Image-based size estimation of broccoli heads under varying degrees of occlusion. *Biosyst. Eng.* **2021**, *208*, 213–233. [CrossRef]
40. Kaskman, R.; Zakharov, S.; Shugurov, I.; Ilic, S. Homebreweddb: RGB-D dataset for 6D pose estimation of 3D objects. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, Seoul, Republic of Korea, 27–28 October 2019.
41. Koch, T.; Liebel, L.; Fraundorfer, F.; Korner, M. Evaluation of cnn-based single-image depth estimation methods. In Proceedings of the European Conference on Computer Vision (ECCV) Workshops, Munich, Germany, 8–14 September 2018.
42. Tancik, M.; Weber, E.; Ng, E.; Li, R.; Yi, B.; Wang, T.; Kristoffersen, A.; Austin, J.; Salahi, K.; Ahuja, A.; et al. Nerfstudio: A Modular Framework for Neural Radiance Field Development. In Proceedings of the ACM SIGGRAPH 2023 Conference Proceedings, Los Angeles, CA, USA, 6–10 August 2023. [CrossRef]
43. Remondino, F.; Karami, A.; Yan, Z.; Mazzacca, G.; Rigon, S.; Qin, R. A critical analysis of nerf-based 3d reconstruction. *Remote Sens.* **2023**, *15*, 3585. [CrossRef]
44. Jiang, H.; Li, R.; Sun, H.; Tai, Y.W.; Tang, C.K. Registering Neural Radiance Fields as 3D Density Images. *arXiv* **2023**, arXiv:2305.12843.
45. Petrovska, I.; Jäger, M.; Haitz, D.; Jutzi, B. Geometric Accuracy Analysis between Neural Radiance Fields (NeRFs) and Terrestrial laser scanning (TLS). *Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci.* **2023**, *48*, 153–159. [CrossRef]
46. Oechsle, M.; Peng, S.; Geiger, A. Unisurf: Unifying neural implicit surfaces and radiance fields for multi-view reconstruction. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 5589–5599.
47. Jäger, M.; Jutzi, B. 3D Density-Gradient Based Edge Detection on Neural Radiance Fields (NeRFS) for Geometric Reconstruction. *Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci.* **2023**, *48*, 71–78. [CrossRef]
48. Ni, H.; Lin, X.; Ning, X.; Zhang, J. Edge Detection and Feature Line Tracing in 3D-Point Clouds by Analyzing Geometric Properties of Neighborhoods. *Remote Sens.* **2016**, *8*, 710. [CrossRef]
49. Mutneja, V. Methods of Image Edge Detection: A Review. *J. Electr. Electron. Syst.* **2015**, *4*, 5. [CrossRef]
50. Besl, P.J.; McKay, N.D. Method for registration of 3-D shapes. In *Proceedings of the Sensor Fusion IV: Control Paradigms and Data Structures*; SPIE: Bellingham, WA, USA, 1992; Volume 1611, pp. 586–606.
51. Hodson, T.O. Root-mean-square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geosci. Model Dev.* **2022**, *15*, 5481–5487. [CrossRef]
52. Püschel, J. Vergleich eines 3D-Modells zwischen Bundler und Breuckmann. Bachelor's Thesis, Institute for Photogrammetry and Remote Sensing, Karlsruhe Institute of Technology—KIT, Karlsruhe, Germany, 2011.
53. Jäger, M.; Landgraf, S.; Jutzi, B. Density Uncertainty Quantification with NeRF-Ensembles: Impact of Data and Scene Constraints. *arXiv* **2023**, arXiv:2312.14664.