Article

A Critical Analysis of NeRF-Based 3D Reconstruction

Fabio Remondino [1,*] , Ali Karami [1] , Ziyang Yan [1,2], Gabriele Mazzacca [1,3], Simone Rigon [1] and Rongjun Qin [4,5,6,7]

1   3D Optical Metrology Unit, Bruno Kessler Foundation (FBK), Via Sommarive 18, 38123 Trento, Italy;
    akarami@fbk.eu (A.K.); zyan@fbk.eu (Z.Y.); gmazzacca@fbk.eu (G.M.); srigon@fbk.eu (S.R.)
2   Department of Information Engineering and Computer Science, University of Trento, 38123 Trento, Italy
3   Department Mathematics, Computer Science and Physics, University of Udine, 33100 Udine, Italy
4   Geospatial Data Analytics Laboratory, The Ohio State University, 218B Bolz Hall, 2036 Neil Avenue,
    Columbus, OH 43210, USA; qin.324@osu.edu
5   Department of Civil, Environmental and Geodetic Engineering, The Ohio State University, 218B Bolz Hall,
    2036 Neil Avenue, Columbus, OH 43210, USA
6   Department of Electrical and Computer Engineering, The Ohio State University, 2036 Neil Avenue,
    Columbus, OH 43210, USA
7   Translational Data Analytics Institute, The Ohio State University, 1760 Neil Avenue,
    Columbus, OH 43210, USA
*   Correspondence: remondino@fbk.eu

Abstract: This paper presents a critical analysis of image-based 3D reconstruction using neural radiance fields (NeRFs), with a focus on quantitative comparisons with respect to traditional photogrammetry. The aim is, therefore, to objectively evaluate the strengths and weaknesses of NeRFs and provide insights into their applicability to different real-life scenarios, from small objects to heritage and industrial scenes. After a comprehensive overview of photogrammetry and NeRF methods, highlighting their respective advantages and disadvantages, various NeRF methods are compared using diverse objects with varying sizes and surface characteristics, including textureless, metallic, translucent, and transparent surfaces. We evaluated the quality of the resulting 3D reconstructions using multiple criteria, such as noise level, geometric accuracy, and the number of required images (i.e., image baselines). The results show that NeRFs exhibit superior performance over photogrammetry in terms of non-collaborative objects with texture-less, reflective, and refractive surfaces. Conversely, photogrammetry outperforms NeRFs in cases where the object's surface possesses cooperative texture. Such complementarity should be further exploited in future works.

Keywords: photogrammetry; neural radiance fields; NeRF; 3D reconstruction; quality; accuracy

1. Introduction

In the fields of computer vision and photogrammetry, high-quality 3D reconstruction is an important topic that has many applications, such as quality inspection, reverse engineering, structural monitoring, digital preservation, etc. However, low-cost, portable, and flexible 3D measuring techniques that provide high geometric accuracy and high-resolution details have been in great demand for some years. Existing methods for 3D reconstruction can be broadly categorized as either contact or non-contact techniques [1]. In order to determine the precise 3D shape of an object, contact-based techniques often employ physical tools like a caliper or a coordinate measurement machine. While precise geometrical 3D measurements are feasible and well-suited for many applications, they do have some drawbacks, such as the length of time required to acquire data and perform sparse 3D reconstruction, the limitations of the measuring system, and/or the need for expensive instrumentation, which limits their use to specialized laboratories and projects with unique metrological specifications. Non-contact technologies, on the other hand, allow for accurate 3D reconstruction without the associated drawbacks. Most researchers

have focused on passive image-based approaches due to their low cost, portability and flexibility over a wide range of application fields, including industrial inspection and quality control [2–5] as well as heritage 3D documentation [6–9].

Among image-based 3D reconstruction approaches, **photogrammetry** is a widely recognized method that can create a dense and geometrically accurate 3D point cloud of a real-world scene from a set of images taken from different perspectives. Photogrammetry can handle a wide range of scenes, from indoor to outdoor environments and has a proven track record in multiple projects with many commercial and open-source tools available [10,11]. However, photogrammetry has its limitations, particularly when it comes to the 3D measurement of non-collaborative surfaces due to its sensitivity to textural properties of objects, and it can struggle with the generation of highly detailed 3D reconstructions. The presence of specular reflections in images, for instance, can result in noisy results for highly reflective and weakly textured objects, while transparent objects can pose significant challenges due to texture changes induced by refraction and mirror-like reflections [12–15].

More recently, a novel approach for the 3D reconstruction from image datasets based on **Neural Radiance Fields (NeRFs)** has attracted significant attention in the research community [16–22]. This approach is capable of producing novel views of complex scenes by optimizing a continuous scene function from a set of oriented images. NeRF works by training a fully connected network, referred to as a neural radiance field, to replicate the input views of a scene through the use of a rendering loss (Figure 1).
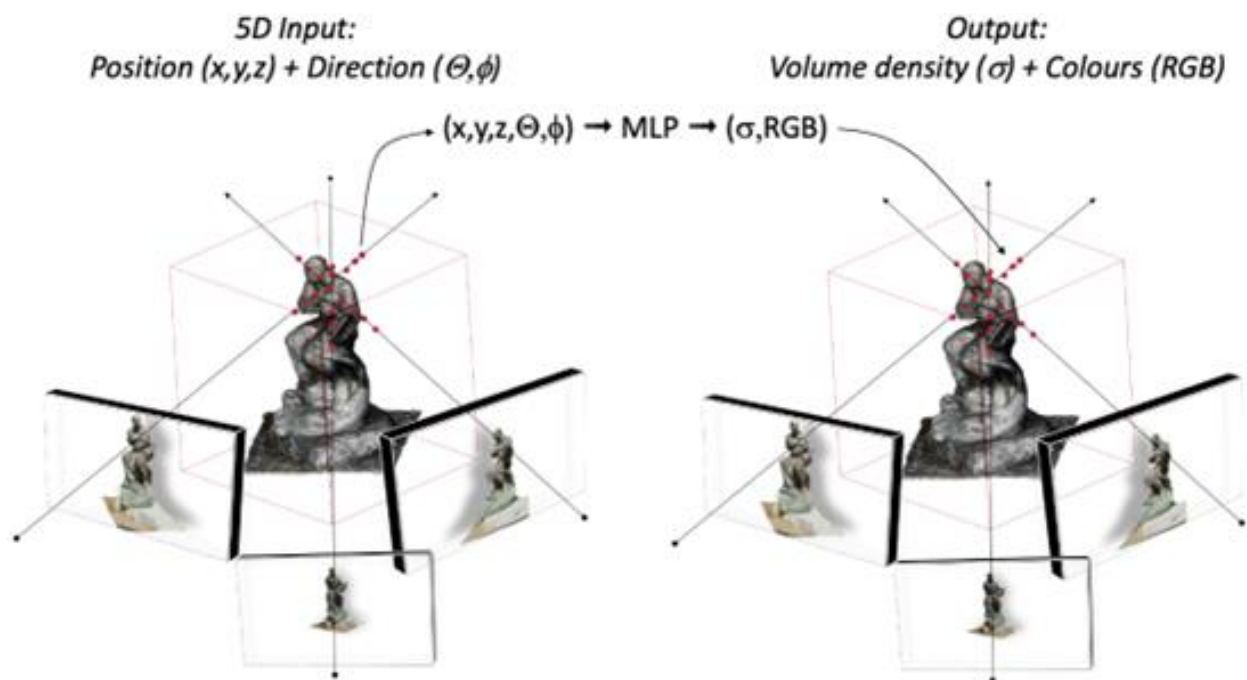

Figure 1. The basic concept of NeRF scene representation (after [16])—see also Section 2.2.

As shown in Figure 1, the neural network takes as input a set of continuous 5D coordinates consisting of spatial locations (x, y, z) and viewing directions (θ, ϕ), and it outputs the volume density (σ) and view-dependent emitted radiance (RGB) in each direction at each point. The NeRF is then rendered from a certain perspective, and 3D geometry can be derived, e.g., in the form of a mesh by marching camera rays [23].

Despite their recent popularity, however, there remains a need for a critical analysis of NeRF-based methods in comparison to the more conventional photogrammetry in order to objectively quantify the quality of resulting 3D models and to fully understand their strengths and limitations.

Aims of This Research

NeRF methods have recently emerged as a promising alternative to photogrammetry and computer vision in the field of image-based 3D reconstruction. Therefore, this research seeks to thoroughly analyze NeRF approaches for 3D reconstruction purposes. We evaluate the accuracy of the 3D reconstruction generated using NeRF-based techniques and via photogrammetry on a wide variety of objects ranging in size and surface characteristics (well-textured, texture-less, metallic, translucent and transparent). We examined the data generated by each technique in terms of surface deviation (noise level) and geometric accuracy. The final aim is to assess the applicability of the NeRF method in a real-world scenario and to provide objective evaluation metrics regarding the advantages and limitations of NeRF-based 3D reconstruction approaches.

The paper is organised as follows: an overview of previous research activities for 3D reconstruction using both photogrammetric-based and NeRF-based approaches is presented in Section 2. Section 3 presents the proposed quality evaluation pipeline and employed datasets, whereas Section 4 reports the evaluation and comparison results. Finally, conclusions and future research plans are provided in Section 5.

2. The State of the Art

In this section, a comprehensive overview of previous research on 3D reconstruction is conducted, incorporating both photogrammetric and NeRF-based approaches and considering their application to non-collaborative surfaces (reflective, textureless, etc.).

2.1. Photogrammetric-Based Methods

Photogrammetry is a widely accepted method for 3D modeling of well-textured objects and is capable of accurately and reliably recovering the 3D shape of an object through multi-view stereo (MVS) methods. Photogrammetric-based methods [19,24–30] either rely on feature matching for depth estimation [27,28] or use voxels to represent shapes [24,29,31,32]. Learning-based MVS methods can also be used, but they typically replace certain parts of the classic MVS pipeline, such as feature matching [33–36], depth fusion [37,38], or multi-view image depth inference [39–41]. However, objects with texture-less, reflective, or refractive surfaces are challenging to reconstruct because all photogrammetric methods require matching correspondences across multiple images [14]. To address this, various photogrammetric methods have been developed to reconstruct these non-collaborative objects. For texture-less objects, solutions such as random pattern projection [13,42,43] or synthetic pattern [14,44] have been suggested. However, these methods struggle with highly reflective surfaces with strong specular reflections or interreflection [43]. Other methods like cross polarisation [7,45] and image pre-processing [46,47] have been used for reflective or non-collaborative surfaces, but some techniques can potentially smooth off surface roughness and affect texture consistency across views [48,49]. Photogrammetry is also utilized in hybrid methods [50–53], where MVS approaches are used to generate a sparse 3D shape that can serve as a base for high-resolution measurements using Photometric Stereo (PS). Conventional [52,54,55] and learning-based [56–58] PS methods are also used to understand the image irradiance equation and retrieve the geometry of the imaged object but specular surfaces are still challenging for all image-based methods.

2.2. NeRF-Based Methods

Synthesizing photo-realistic images and videos is at the heart of computer graphics and has been the focus of decades of research [59]. Neural rendering is a learning-based class of image and video generation approach to control scene properties (e.g., illumination, camera parameters, pose, geometry, appearance, etc.). Neural rendering combines deep learning methods with physical knowledge from computer graphics to achieve controllable and photo-realistic (3D) models of scenes. Among them, NeRF, first proposed by Mildenhall et al. in 2020, is a method for rendering new views and reconstructing 3D scenes using an implicit representation (Figure 1). In the NeRF approach, a neural network is employed to

learn the 3D shape of an object from 2D images. The radiance field, as defined in Equation (1), captures the color and volume density for each point in the scene from every possible viewing direction:

$$F(X, d) ====> (c, \sigma) \tag{1}$$

The NeRF model utilizes a neural network representation where X represents the 3D coordinate of the images, d represents the azimuthal and polar viewing angles, c represents color and σ represents the volume density of the scene. In order to ensure multi-view consistency, the prediction of σ is designed to be independent of the viewing direction, while the color c can vary based on both the viewing direction and position. To achieve this, a Multi-Layer Perceptron (MLP) is employed in two steps. In the first step, the MLP takes X as input and outputs both σ and a high-dimensional feature vector. The feature vector is then combined with the viewing direction d and passed through an additional MLP, which produces the color representation c.

The original NeRF implementation, as well as subsequent methods, utilized a non-deterministic stratified sampling approach, which is described by Equations (2)–(4). This method involved dividing the ray into N equally spaced bins and uniformly drawing a sample from each bin:

$$C(r) = \sum_{i=1}^{N} \alpha_i T_i c_i \tag{2}$$

$$T_i = e^{-\sum_{j=1}^{i-1} \sigma_j \delta_j} \tag{3}$$

$$\alpha_i = 1 - e^{\sigma_i \delta_i} \tag{4}$$

where δ_i denotes the distance between the consecutive samples (i and $i + 1$), while σ_i and c_i represent the estimated density and color values along the sample point (i). The transparency or opacity α_i at sample point (i) is computed also using Equation (4).

Successive methods [60–62] have also incorporated the estimated depth, as expressed in Equation (5), to impose restrictions on densities, making them resemble delta-like functions at the surfaces of the scene, or to enforce smoothness in depth:

$$D(r) = \sum_{i=1}^{N} \alpha_i t_i T_i \tag{5}$$

To optimize the MLP parameters, a square error photometric loss is used for each pixel:

$$M = \sum_{r \in R} \left\| C - C_{gt} \right\|_2^2 \tag{6}$$

where the variable $C_{gt}(R)$ represents the ground truth color of the pixel in the training image that corresponds to the ray r, while R refers to the batch of rays associated with the image to be synthesized. It should be noted that the learned implicit 3D representation of an NeRF is designated for view rendering. To obtain the explicit 3D geometry, depth maps for different views need to be extracted by taking the maximal likelihood of depth distribution for each ray. These depth maps can then be fused to derive point clouds or fed into the Marching Cube [23] algorithm to derive 3D meshes.

Although NeRF provides an alternative solution for 3D reconstruction compared to traditional photogrammetry methods and can produce promising results in situations where photogrammetry may fail to deliver accurate results, it still faces several limitations, as reported by different authors [63–68]. Some of the main issues from a 3D metrological perspective that need to be considered include:

(1)   The resolution of the generated neural renderings (afterward converted into a 3D mesh) can be limited by the quality and resolution of the input data. In general, higher-resolution input data will result in a higher-resolution 3D mesh, but the tradeoff is increased computational requirements.

(2)   Generating a neural rendering (and then a 3D mesh) using NeRF can be computationally intensive, requiring significant amounts of computing power and memory.

(3)   The general inability to accurately model the 3D shape of non-rigid objects.

(4)   The original NeRF model is optimized based on a per-pixel RGB reconstruction loss, which can result in a noisy reconstruction as an infinite number of photo-consistent explanations exist when using only RGB images as input.

(5)   NeRF generally requires a large number of input images with small baselines to generate an accurate 3D mesh, especially for scenes with complex geometry or occlusions. This can be a challenge in situations where images are difficult to acquire or when computational resources are limited.

To face the above issues, researchers have proposed several modifications and extensions to the original NeRF method in order to improve performance and 3D results. Tancik et al. [69] and Sitzmann et al. [70] adopted the position encoding operation with a different frequency to NeRFs in order to improve the resolution of the neural rendering outcome since high-frequency representation capacity in NeRFs is insufficient. Following this, other approaches have focused on improving the efficiency and resolution of the neural rendering outcome in different ways, including model acceleration [20,71], compression [72–74], relighting [75–77], and View-Dependence Normalization [78] (Zhu et al., 2023), or high-resolution 2D feature planes [68]. Müller et al. [20] introduced the concept of instant Neural Graphics Primitives with a Multiresolution Hash Encoding, which allows for fast and efficient generation of 3D models. Barron et al. [64,79] proposed that Mip-NeRF, a modified version of the original NeRF, allows for the representation of scenes on continuously valued scales. Mip-NeRF greatly increases the capacity of NeRF to emphasize fine details by efficiently rendering anti-aliased conical frustums instead of rays. However, limitations of the method may include the difficulty in training and issues with computational efficiency. Chen et al. [72] presented a new method called Tensorf for modeling and reconstructing the radiance fields of a scene as a 4D tensor. This approach represents a 3D voxel grid with per-voxel multi-channel features. In addition to providing superior rendering quality, this method achieves much lower memory usage compared to previous and contemporary methods. Yang et al. [80] presented a fusion-based approach called PS-NeRF that combines the strengths of NeRF with photometric stereo methods. This method aims to address the limitations of traditional photometric stereo techniques by utilizing NeRF's capability to reconstruct a scene, ultimately leading to an improved resolution of the resultant mesh. Reiser et al. [68] introduced Memory-Efficient Radiance Field (MERF) representation, which allows for the fast rendering of large-scale scenes by utilizing a sparse feature grid and high-resolution 2D feature planes. Li et al. [21] introduced Neuralangelo, an innovative method that utilizes multi-resolution 3D hash grids and neural surface rendering to achieve superior results in recovering dense 3D surface structures from multi-view images, enabling highly detailed large-scale scene reconstruction from RGB video captures.

Some approaches [67,81–85] have been proposed that extend NeRF to a dynamic domain. These approaches make it possible to reconstruct and render images of objects while they are undergoing rigid and non-rigid motions from a single camera that is moving around the scene. For example, Yan et al. [84] introduced a surface-aware dynamic NeRF (NeRF-DS) and a mask-guided deformation field. By incorporating surface position and orientation as conditioning factors in the neural radiance field function, NeRF-DS improves the representation of complex reflectance properties in specular surfaces. Additionally, the use of a mask-guided deformation field enables NeRF-DS to effectively handle large deformations and occlusions occurring during object motion.

To improve the accuracy of 3D reconstruction in the presence of noise, particularly for smooth and texture-less surfaces, some studies incorporated various priors into the

optimization process. These priors include semantic similarity [86], depth smoothness [60], surface smoothness [87,88], Manhattan world assumptions [89], and monocular geometric priors [90]. In contrast, the NoPe-NeRF method proposed by Bian et al. [91] uses mono-depth maps to constrain the relative poses between frames and regularize NeRF's geometry. This method results in better pose estimation, which improves the quality of novel view synthesis and geometry reconstruction. Rakotosaona et al. [92] introduced a novel and versatile architecture for 3D surface reconstruction, which efficiently distills volumetric representations from NeRF-driven approaches into a Signed Surface Approximation Network. This approach enables the extraction of accurate 3D meshes and appearance while maintaining real-time rendering capabilities across various devices. Elsner et al. [93] presented Adaptive Voronoi NeRFs, a technique that enhances the efficiency of the process by employing Voronoi diagrams to partition the scene into cells. These cells are subsequently subdivided to effectively capture and represent intricate details, leading to improved performance and accuracy. Similarly, Kulhanek and Sattler [94] introduced a new radiance field representation called tera-NeRF, which successfully adjusts to 3D geometry priors given as a sparse point cloud for exploiting more details. However, it is worth noting that the quality of rendered scenes may differ depending on the density of the point cloud in various regions.

Some works aimed to reduce the number of input images [60,70,78,86,90,95]. Yu et al. [95] presented an architecture that conditions NeRF on image inputs using a fully convolutional method, enabling the network to learn a scene prior to being trained on multiple scenes. This allows it to perform feed-forward view synthesis from a small number of viewpoints, even as few as one. Similarly, Niemeyer et al. [60] introduced a method to sample unseen views and regularize the appearance and geometry of patches generated from these views. Jain et al. [86] proposed DietNeRF to enhance few-shot quality via an auxiliary semantic consistency loss that boosts realistic renderings of new positions. DietNeRF learns from individual scenes to accurately render input images from the same position and to match high-level semantic features across diverse, random poses.

In the field of cultural heritage, only a limited number of publications have explicitly investigated and recognized the potential of NeRFs for 3D reconstruction, digital preservation and conservation purposes [96,97].

3. Analysis and Evaluation Methodology

The main goal is to conduct a critical evaluation of NeRF-based methods with respect to conventional photogrammetry by objectively measuring the quality of resulting 3D data. To accomplish this, a variety of objects and scenes with different sizes and surface characteristics, including well-textured, texture-less, metallic, translucent, and transparent, are considered (Section 3.3). The proposed evaluation strategy and metrics (Sections 3.1 and 3.2) should help researchers to understand the strengths and limitations of each approach and could be adopted for quantitative evaluations of newly proposed methods. All experiments are based on the SDFStudio [98] and Nerfstudio [22] frameworks. It is worth reminding that the NeRF output is a neural rendering; therefore, a marching cube approach [23] is used to create a mesh model from the different depth maps of each view. A point cloud is then extracted from the mesh vertices for the quantitative evaluations using the Open3D library [78].

3.1. Proposed Methodology

Firstly, various NeRF methods available in dedicated frameworks [22,98] are applied to two datasets in order to understand their performances and choose the most outperforming method (Section 4.1). Then, this method is applied to other datasets to run evaluation and comparisons (Sections 4.2–4.7) with respect to conventional photogrammetry and the available ground truth (GT) data.

Figure 2 shows the general overview of the proposed procedure to quantitatively assess the performance of an NeRF-based 3D reconstruction. All collected images or videos

require camera poses in order to generate 3D reconstructions, either with conventional photogrammetry or NeRF-based methods. Starting from the available images, camera poses are retrieved using Colmap. Then, a multi-view stereo (MVS) or NeRF is applied to generate 3D data. Finally, we provide a unique and robust environment and conditions to provide an objective geometric comparison. To achieve this, 3D data produced with photogrammetry and NeRF are co-registered and rescaled with respect to the available ground truth (GT) data in Cloud Compare (using an Iterative Closest Point (ICP) algorithm [99] and a quality evaluation is performed. To provide an unbiased evaluation of geometric accuracy, different well-known criteria are applied [13,43,100–102], including best plane fitting, cloud-to-cloud comparison, profiling, accuracy, and completeness. For the first two criteria, metrics, such as Standard Deviation (*STD*), Mean Error (*Mean_E*), Root Mean Squares Error (*RMSE*) and Mean Absolute Error (*MAE*), are used (Section 3.2).
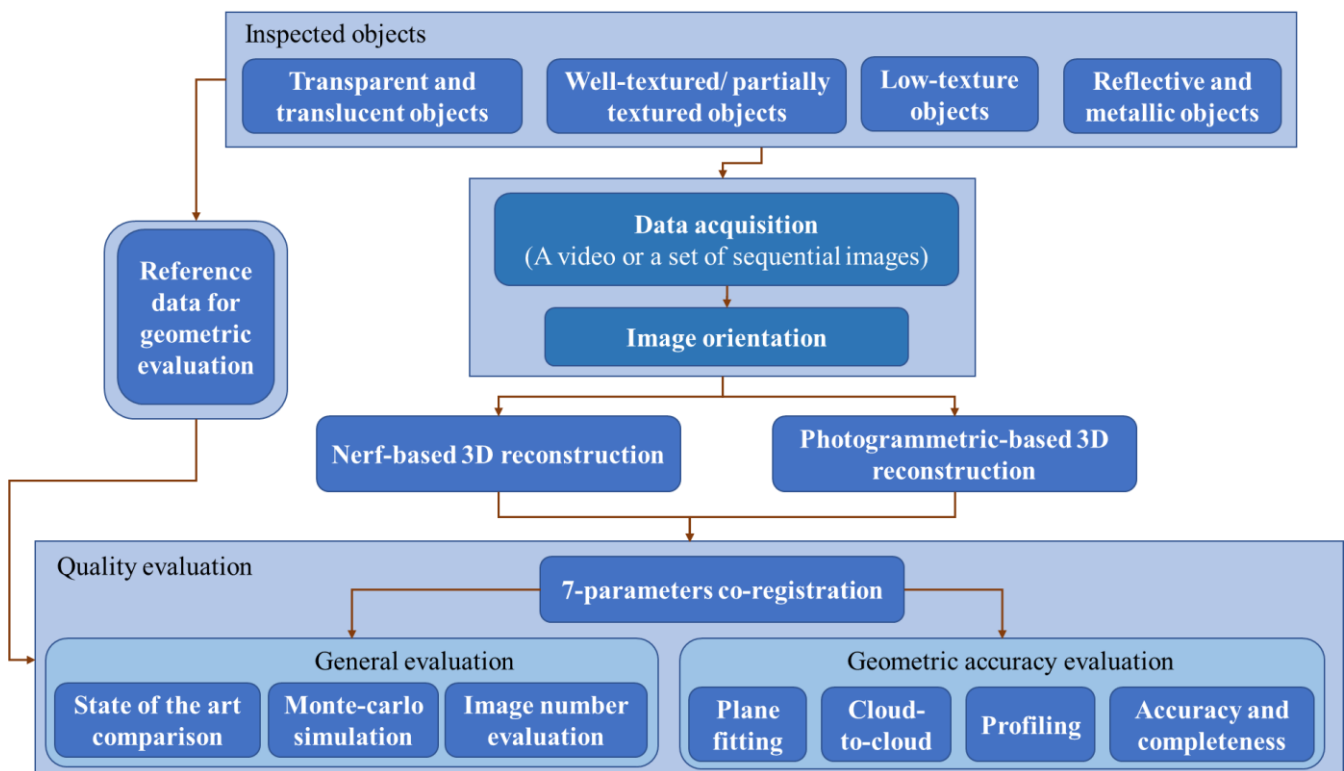

Figure 2. Overview of the proposed procedure to assess the performance of NeRF-based 3D reconstruction with respect to conventional photogrammetry.

Best plane fitting is accomplished by using a Least Squares Fitting (LSF) algorithm that defines a best-fitted plane on an area of the object, which is assumed to be planar. This criterion allows us to evaluate the level of noise in the 3D data generated by photogrammetry or NeRF methods.

Profiling is carried out by extracting a cross-section from the 3D data to highlight complex geometric details of the reconstructed surface. An inspection of profiles allows us to evaluate the performance of a method in preserving geometric details, such as edges and corners, and avoid smoothing effects.

Cloud-to-cloud (C2C) comparison refers to the measurement of the nearest neighboring distance between corresponding points in two point clouds.

3.2. Metrics

Despite the increasing popularity and widespread application of NeRF to 3D reconstruction purposes, there is still a shortage of information on quality assessment based on a specified standard or criteria (e.g., the VDI/VDE 2643 BLATT 3). Following the

co-registration process and criteria mentioned before, the following metrics are used (in particular for cloud-to-cloud and plane fitting processes):

$$STD = \sqrt{\frac{1}{N-1}\sum_{j=1}^{N}(X_j - \underline{X})^2} \qquad (7)$$

$$Mean\_E = \frac{(X_1 + X_2 + \cdots X_j)}{N} \qquad (8)$$

$$RMSE = \sqrt{\frac{\sum_{j=1}^{N}(X_j)^2}{N}} \qquad (9)$$

$$MAE = \sqrt{\frac{\sum_{j=1}^{N}|X_j|}{N}} \qquad (10)$$

where N denotes the number of observed point clouds, X_j denotes the closest distance of each point to the corresponding reference point or surface, and X denotes the average observed distance.

Accuracy and completeness, respectively, also known as precision and recall [101,102], involve measuring the distance between two models. When assessing accuracy, the distance is computed from the computed data to a ground truth (GT). Conversely, to evaluate completeness, the distance is computed from the GT to the computed data. These distances can be either signed or unsigned, depending on the specific evaluation method. Accuracy reflects how closely the reconstructed points align with the ground truth, while completeness indicates the degree to which all GT points are covered. Typically, a threshold distance is employed to determine the fraction or percentage of points that fall within the acceptable threshold. The threshold value is determined based on factors such as data density and noise levels.

3.3. Testing Objects

To achieve the work objectives, different datasets are used (Figure 3): they feature objects of different dimensions and surface types, and they were captured under different lighting conditions, materials, camera networks, scales, and resolutions.
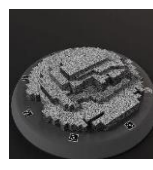

	Ignatius	Truck	Stair	Synthetic	Industrial	Bottle_1	Bottle_2
numb. images	260 images	235 images	200 images	200 images	140 images	360 images	360 images
resolution	1920 × 1080 px	1920 × 1080 px	1920 × 1080 px	1920 × 1080 px	1080 × 1920 px	1080 × 1920 px	1080 × 1920 px
ground truth (GT)	laser scanner	laser scanner	ideal planes	CAD model	laser scanner	powdered photogrammetry	powdered photogrammetry
characteristics	outdoor/partially textured	outdoor/partially textured/reflective	indoor/well-texture/reflective	small/complex/well-texture	complex/textureless/highly reflective	transparent/refractive	translucent/semi-transparent/refractive

Figure 3. Set of objects, with different surface characteristics, used to evaluate NeRF methods.

The Ignatius and Truck datasets are derived from the Tanks and Temples benchmark [101], where GT data (acquired with laser scanning) are also available.

The other datasets (Stair, Synthetic, Industrial, Bottle_1 and Bottle_2) are created in FBK. The Stair dataset offers a flat, reflective, and well-textured surface with sharp edges. GT is provided by ideal plans of the step surfaces. The Synthetic 3D object created using Blender v3.2.2 (for the geometric model, UV texture and material) and Quixel Mixer v2022 (for PBR textures) has a well-textured surface featuring complex geometry, including edges and corners. A virtual camera with specific parameters (Focal length: 50 mm; Sensor size: 36 mm; image size: 1920 × 1080 pixels) is used to create a sequence of images that follows a spiral curvy path around the object. The 3D model generated in Blender is used as GT for the accuracy assessment. The Industrial object has a textureless and highly reflective metallic surface which raises problems for all passive 3D methods. Its GT data are acquired with a Hexagon/AICON Primescan active scanner with a nominal accuracy of 63 μm. Two bottles are also included, featuring transparent and refractive surfaces: their GT data are generated using photogrammetry after powdering/spraying the surfaces.

A specific benchmark for NeRF methods is under preparation by the authors and will be available at https://github.com/3DOM-FBK/NeRFBK [103], containing many more datasets with ground truth data.

4. Comparisons and Analyses

This section presents experiments that evaluate and compare the performance of NeRF-based techniques versus standard photogrammetry (Colmap). After comparing multiple state-of-the-art methods (Section 4.1), Instant-NGP was selected as the NeRF-based method to be fully assessed, as it delivered superior results with respect to the other methods. The NeRF training was executed using a Nvidia A40 GPU, while the geometric comparisons of the 3D results were performed on a standard PC.
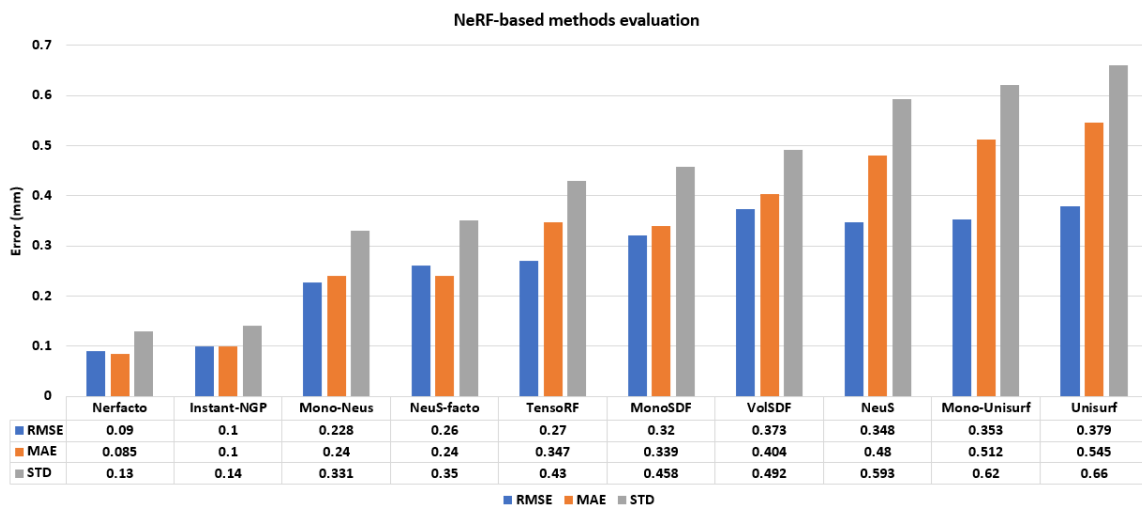
4.1. State-of-the-Art Comparison

The primary objective is to conduct a comprehensive analysis of multiple NeRF-based methods. To achieve this goal, the SDFStudio unified framework developed by Yu et al. [98] is used as it incorporates multiple neural implicit surface reconstruction approaches into a single framework. SDFStudio is built upon the Nerfstudio framework [22]. Among the implemented approaches, ten were chosen in order to compare their performances: Nerfacto and Tensorf from Nerfstudio, Mono-Neus, Neus-Facto, MonoSDF, VolSDF, NeuS, Mono-Unisurf and UniSurf from SDFStudio and InstantNGP from its original implementation in Müller et al. [20].

Two datasets are used: (i) the Synthetic dataset, composed of 200 images (1920 × 1080 px) and (ii) the Ignatius dataset [101], which contains 263 images (extracted from a video at 1920 × 1080 px resolution).

The comparison results with respect to GT data are reported in Figure 4. Results in terms of RMSE, MAE, and STD show that Instant-NGP and Nerfacto methods achieved the best outcomes, outperforming all other methods. In terms of processing time, Instant-NGP required less than a minute for both datasets to train the model, Nerfacto some 15 min. It should be noted that for the Ignatius sequence (Figure 4b), despite the neural rendering for MonoSDF, VolSDF, and Neus-facto being visually satisfactory, the marching cube to export a mesh model failed; hence, no evaluation was possible.

Therefore, based on the achieved accuracies and processing time, Instant-NGP was chosen and employed for the successive experiments in this paper.

NeRF-based methods evaluation

	Nerfacto	Instant-NGP	Mono-Neus	NeuS-facto	TensoRF	MonoSDF	VolSDF	NeuS	Mono-Unisurf	Unisurf
RMSE	0.09	0.1	0.228	0.26	0.27	0.32	0.373	0.348	0.353	0.379
MAE	0.085	0.1	0.24	0.24	0.347	0.339	0.404	0.48	0.512	0.545
STD	0.13	0.14	0.331	0.35	0.43	0.458	0.492	0.593	0.62	0.66

(**a**)

NeRF-based methods evaluation

	Instant-NGP	Nerfacto	TensoRF	Mono-Unisurf	Unisurf	Mono-Neus	NeuS	MonoSDF	VolSDF	neus-facto
RMSE	0.65	0.983	1.293	1.332	1.332	1.917	1.92			
MAE	0.99	1.048	1.842	1.644	1.645	2.624	2.625	Failed	Failed	Failed
STD	0.59	1.435	2.093	2.102	2.103	3.005	3.015			

(**b**)

Figure 4. The comparison results of the various NeRF-based methods on the Synthetic (**a**) and Ignatius (**b**) datasets with 200 and 263 images, respectively.

4.2. Image Baseline's Evaluation

This section reports the assessment of NeRF-based methods when the number of input images is reduced (i.e., the baseline increases). A comparative evaluation between Instant-NGP, identified as the superior method among others (Section 4.1), and Mono-Neus, a well-established approach for sparse image scenarios [66,90], is performed. The experiment utilizes the Synthetic dataset consisting of four subsets of input images, ranging from 200 to 20 images (Figure 5), progressively reducing the number of input images (i.e., approximately doubling the image baselines). For every set of input images, both NeRF methods are used to generate 3D results, keeping a similar number of epochs. For each subset, the RMSE through point-to-point comparison with the GT data is estimated as reported in Figure 5.

The findings depict that Instant-NGP exhibits superior performance compared to Mono-Neus when a large number of input images is available. However, Mono-Neus outperforms Instant-NGP in scenarios where the number of images is low. Nevertheless, it is important to note that neither Instant-NGP nor Mono-Neus are able to successfully generate a 3D reconstruction using only 10 input images.
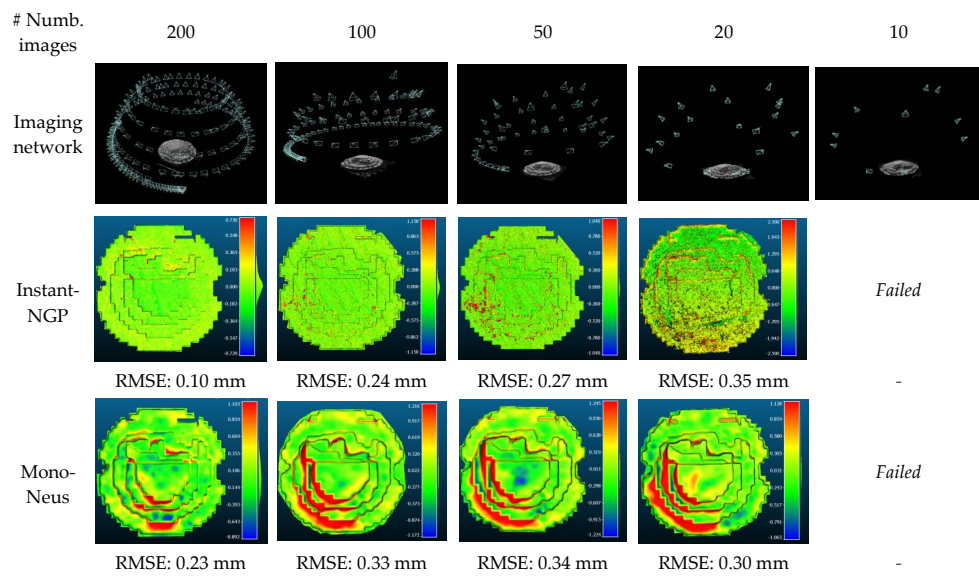
Figure 5. Comparative performance evaluation of Instant-NGP and Mono-Neus on subsets of the Synthetic dataset.

4.3. Monte Carlo Simulation

The aim is to evaluate the quality of NeRF-based 3D results when the camera poses are changed/perturbed. Therefore, a Monte Carlo simulation [104] is employed to randomly perturbate the rotation and translation of camera parameters within a limited range. After the perturbation, using Instant-NGP, a 3D reconstruction is generated and compared to reference data. A total of 30 iterations (runs) are performed within two scenarios: (A) the rotation and translation are randomly disturbed in the range of ±20 mm for translation and ±2 degrees for rotation, (B) rotation and translation are randomly disturbed in the range of ±40 mm and ±4 degrees, respectively. The Ignatius dataset is used to run this simulation, and results are reported in Figure 6 and Table 1. The findings clearly show the importance of having accurate camera parameters. In scenario A, on average, the estimated RMSE is 19.72 mm, with an uncertainty of 2.95 mm. In scenario B, the average estimated RMSE stayed almost the same (19.97 mm), whereas the uncertainty doubled (5.87 mm) due to the larger perturbation range.
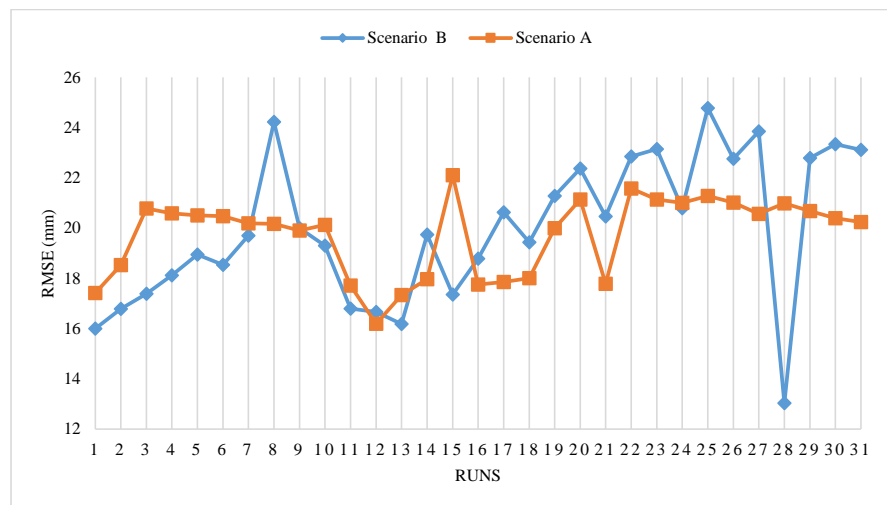

Figure 6. The results of Monte Carlo simulation for perturbing the camera parameters. A summary of the statistics is reported in Table 1.

Table 1. Summary of the Monte Carlo simulation results on the Ignatius dataset. The Error Range is the difference between the Max and Min RMSE, while the Uncertainty is computed as half of the Error Range.

	Scenario A	Scenario B
Perturbing translation (mm)	±20	±40
Perturbing rotation (degree)	±2	±4
Average RMSE (mm)	19.72	19.97
Max RMSE (mm)	22.10	24.78
Min RMSE (mm)	16.90	13.03
Error range (mm)	5.91	11.75
Uncertainty (±mm)	2.95	5.87

4.4. Plane Fitting

A plan-fitting approach can be used to evaluate/measure the level of noise on reconstructed flat surfaces. In the first experiment with the Stair dataset (Figure 7a), photogrammetric point cloud and NeRF-based reconstructions are derived, employing the same number of images and camera poses. Two horizontal planes and three vertical planes are identified and analyzed based on a best-fitting process (Figure 7b). The derived metrics are presented in Table 2.
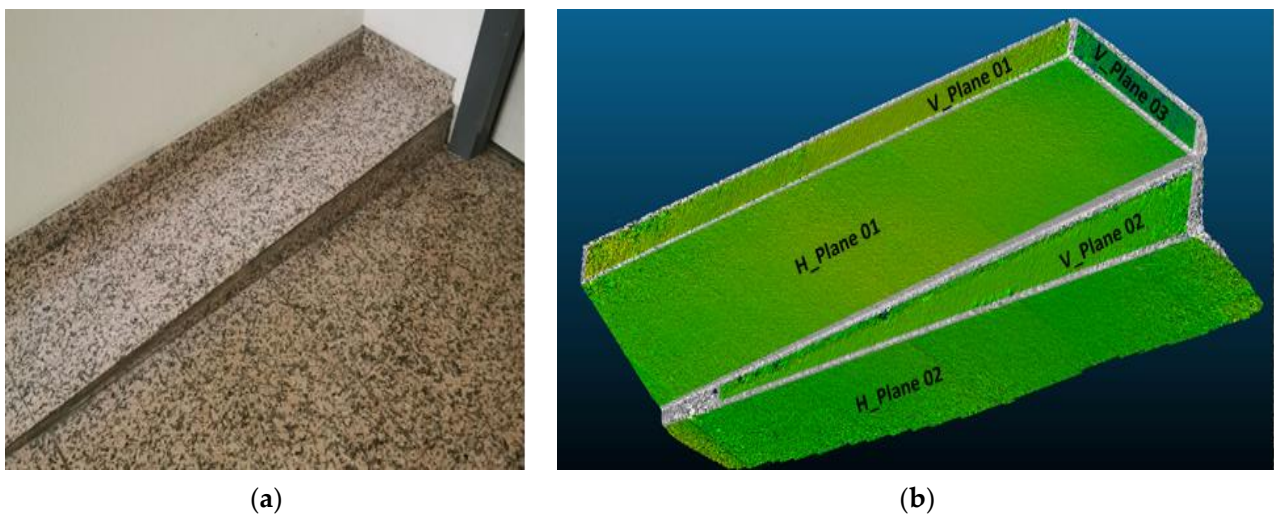

(**a**)   (**b**)

Figure 7. An image of the Step dataset (**a**) and the horizontal and vertical planes used for evaluating the level of the noise in the photogrammetric and NeRF 3D reconstructions (**b**).

Table 2. Evaluating the noise level in the 3D surfaces of the Step dataset processed using photogrammetry and Instant-NGP [Unit: mm].

Plane	Photogrammetry					NeRF				
	RMSE	MAE	STD	Mean Distance	Max Distance	RMSE	MAE	STD	Mean Distance	Max Distance
H_plane_01	0.70	0.61	0.43	0.61	1.65	1.89	1.51	1.38	1.51	6.00
H_plane_02	0.75	0.56	0.55	0.55	2.14	3.53	2.61	2.47	2.60	10.26
V_plane_01	0.78	0.57	0.60	0.57	1.4	2.77	1.79	2.17	1.79	6.69
V_plane_02	0.92	0.77	0.67	0.77	1.94	2.95	2.19	2.06	2.18	6.73
V_plane_03	0.43	0.30	0.37	0.30	1.42	3.10	0.20	2.45	1.96	7.29
Average	0.72	0.56	0.52	0.56	1.65	2.84	1.66	4.74	2.10	7.39

In a similar way, the Synthetic dataset was used, with 200 images for the Instant-NGP and 24 images for the photogrammetric processing. Five vertical and five horizontal planes were selected, as shown in Figure 8, to perform a surface deviation analysis by fitting an ideal plane to the reconstructed object surfaces. Derived metrics are reported in Table 3.
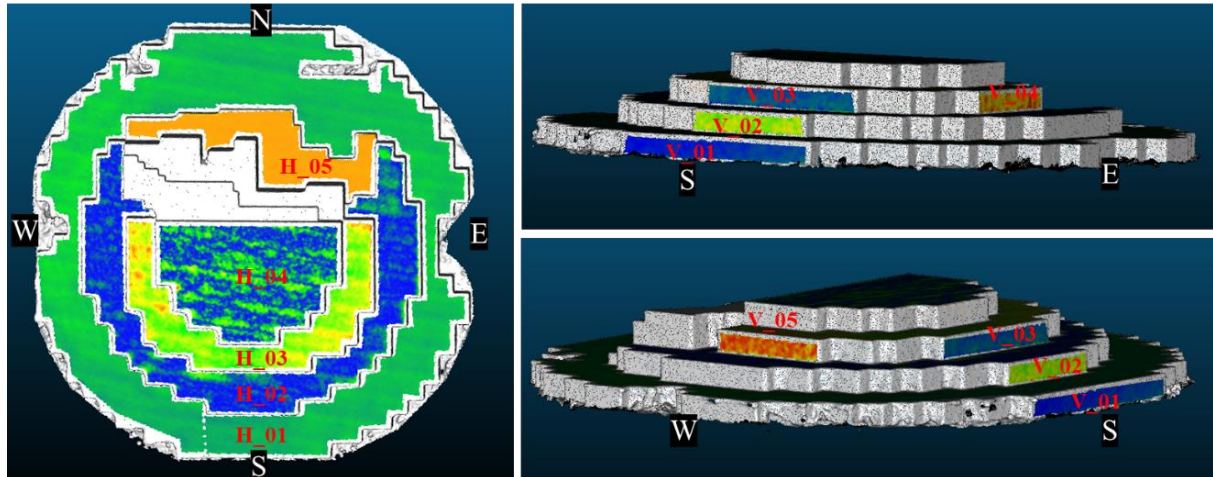

Figure 8. The Synthetic object with some horizontal and vertical planes used for the evaluation.

Table 3. Evaluation metrics for photogrammetry and NeRF results on the Synthetic object [unit: mm].

Plane	Photogrammetry					NeRF				
	RMSE	MAE	STD	Mean Error	Max Error	RMSE	MAE	STD	Mean Error	Max Error
H_plane_01	0.015	0.020	0.017	0.012	0.150	0.070	0.100	0.086	0.070	0.600
H_plane_02	0.014	0.014	0.016	0.010	0.190	0.110	0.120	0.130	0.090	0.600
H_plane_03	0.015	0.011	0.013	0.014	0.200	0.040	0.040	0.060	0.020	0.300
H_plane_04	0.012	0.012	0.013	0.010	0.070	0.040	0.050	0.050	0.040	0.290
H_plane_05	0.080	0.025	0.080	0.020	0.170	0.070	0.080	0.080	0.070	0.450
H_Average	0.027	0.016	0.027	0.013	0.150	0.066	0.078	0.081	0.058	0.448
V_plane_01	0.060	0.046	0.070	0.010	0.450	0.140	0.050	0.150	0.000	0.450
V_plane_02	0.018	0.018	0.020	0.010	0.080	0.050	0.050	0.060	0.030	0.200
V_plane_03	0.039	0.020	0.040	0.014	0.190	0.050	0.070	0.080	0.040	0.25
V_plane_04	0.016	0.016	0.018	0.010	0.090	0.050	0.050	0.070	0.000	0.20
V_plane_05	0.020	0.020	0.027	0.010	0.050	0.070	0.060	0.090	0.000	0.330
V_Average	0.030	0.024	0.035	0.010	0.170	0.072	0.056	0.090	0.014	0.286

From both results (Tables 2 and 3), it is clear that for such two objects, photogrammetry outperforms NeRF, and it can derive less noisy results. NeRF RMSEs are, in general, at least 2–3 times higher than photogrammetry.

4.5. Profiling

The extraction of cross-section profiles is useful to demonstrate the capability of a 3D reconstruction method to retrieve geometric details or apply smoothing effects to the 3D geometry. The results of the Synthetic dataset presented in Section 4.4 are processed using Cloud Compare: several cross-sections are extracted (Figure 9) at predefined distances and geometrically compared against the reference data using different metrics, as reported in Table 4.
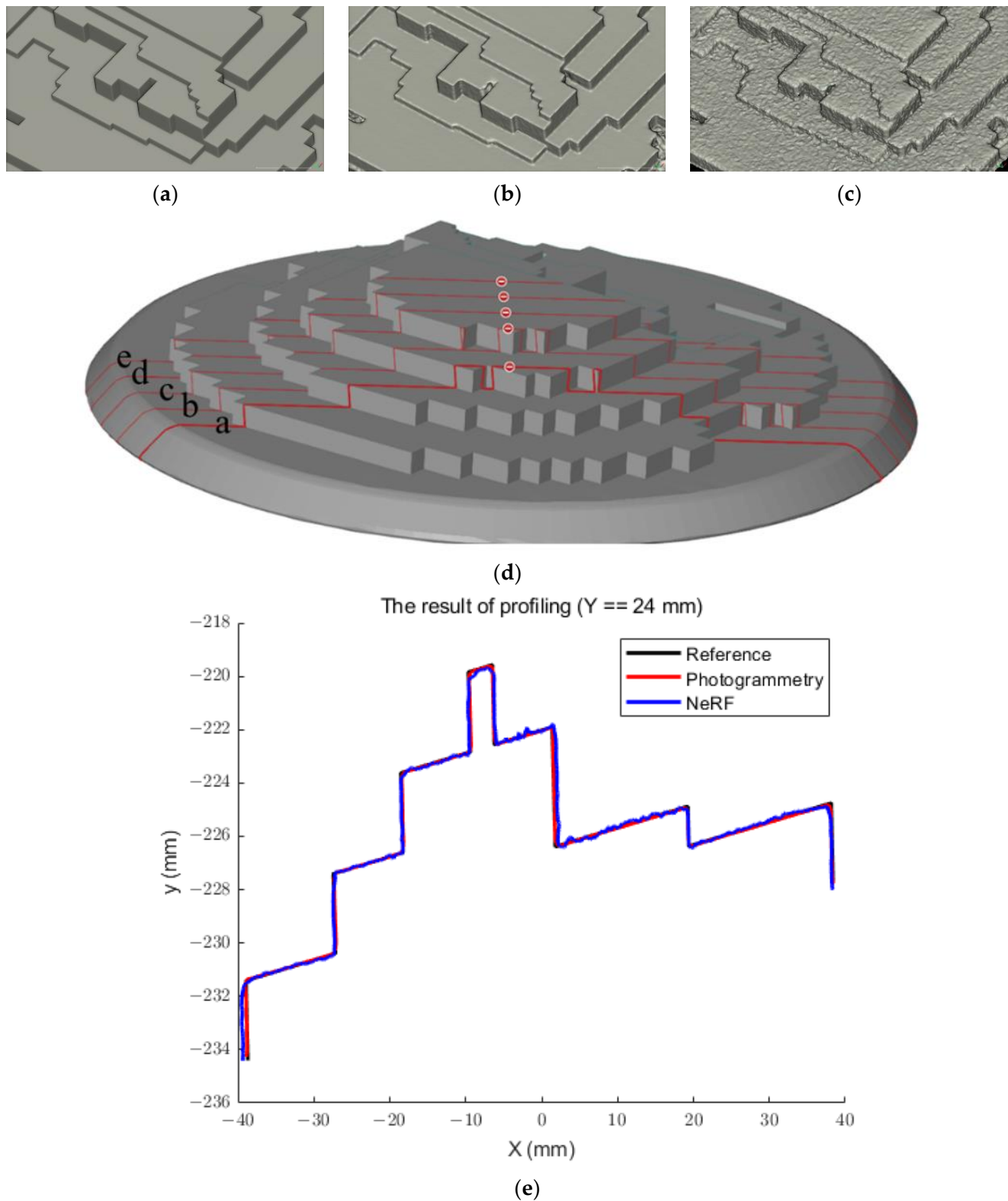
(**a**) (**b**) (**c**)

(**d**)

(**e**)

Figure 9. Close view of the generated meshes for GT (**a**), photogrammetry (**b**) and NeRF (**c**). The different locations of the profiles on the Synthetic object (**d**). An example of a profile on the reference 3D data (black line), photogrammetry (red line) and NeRF (blue line) results (**e**).

Table 4. Comparison of profiles and metrics [unit: mm]—see Figure 9a.

Cross-Section Profiles	Photogrammetry				NeRF			
	RMSE	STD	Mean Error	Max Error	RMSE	STD	Mean Error	Max Error
a	0.09	0.14	0.07	0.20	0.12	0.11	0.04	0.80
b	0.11	0.13	0.08	0.22	0.18	0.15	0.18	1.20
c	0.09	0.05	0.05	0.22	0.16	0.13	0.12	1.40
d	0.08	0.05	0.06	0.21	0.15	0.09	0.09	1.50
e	0.09	0.05	0.07	0.17	0.12	0.21	0.11	1.60
Average	0.09	0.08	0.06	0.20	0.15	0.14	0.11	1.30

The obtained findings for individual cross-sectional profiles, as well as the average of all profiles, show that photogrammetry outperforms NeRF, which generally produces more noisy results (Figure 9a–c). For instance, the average of estimated RMSE and STD for photogrammetry is around 0.09 mm and 0.08 mm, while this value for NeRF is bigger than 0.13 mm.

4.6. Cloud-to-Cloud Comparison

A cloud-to-cloud comparison refers to the assessment of relative Euclidean distances between corresponding 3D samples in a dataset with respect to the reference data. Different objects with different characteristics are considered (Figure 3): Ignatius, Truck, Industrial, and Synthetic. They are small and large-scale objects with texture-less, shiny, and metallic surfaces. For each dataset, 3D data are produced using photogrammetry (Colmap) and Instant-NGP and then co-registered to the available GT (Figure 10). Finally, metrics are derived as reported in Table 5. Worth to notice that the employed number of images is not always the same within the performed tests: indeed, for the Synthetic, Ignatius and Truck datasets, photogrammetry was already providing accurate results with a lower number of images, hence adding more images was not leading to further improvements. On the other hand, for NeRF, all available images were used as fewer images (or an enlargement of the baseline) did not lead to good results (see also Section 4.2).
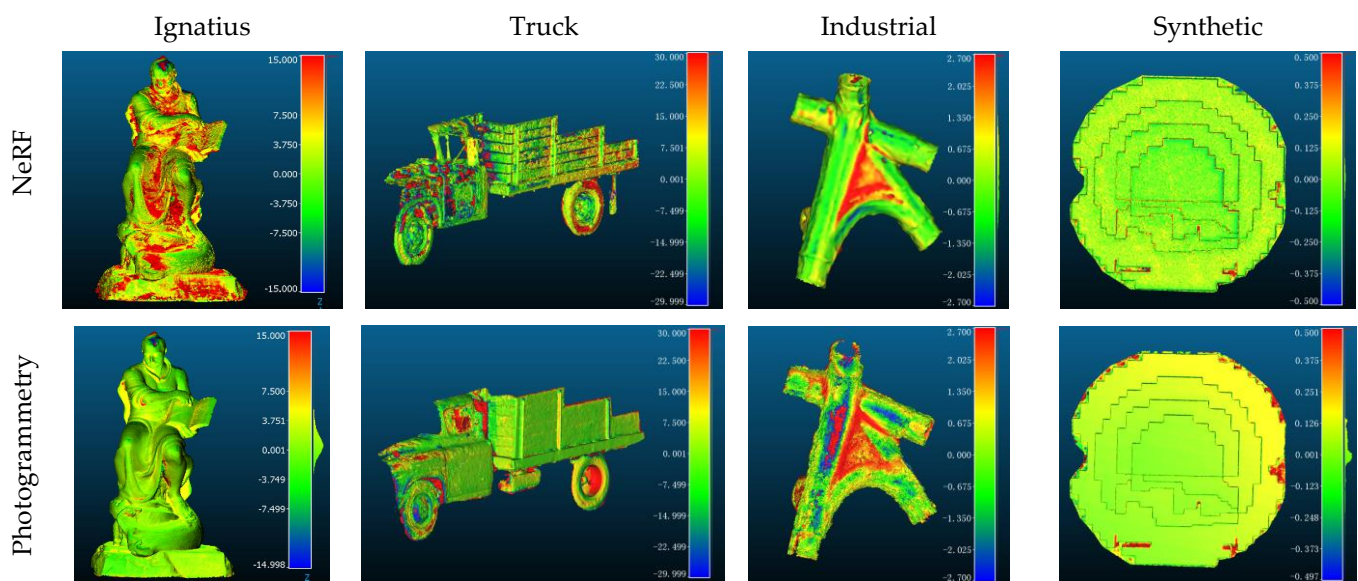

Figure 10. Color-coded cloud-to-cloud comparisons for both Instant-NGP and photogrammetry methods with respect to the ground truth data [unit: mm].

Table 5. Metrics of cloud-to-cloud comparisons for Instant-NGP and photogrammetry methods [unit: mm]. For all but the Industrial object, Photogrammetry was used with a smaller number of images as the achieved accuracy was already better than NeRF.

Object	Method	Images	RMSE	MAE	STD
Industrial	Photogr	140	0.85	1.10	1.50
	NeRF	140	0.78	0.73	1.10
Synthetic	Photogr	25	0.13	0.11	0.17
	NeRF	200	0.19	0.13	0.23
Ignatius	Photogr	130	2.50	0.90	2.00
	NeRF	260	5.50	9.90	4.60
Truck	Photogr	48	16.0	15.0	20.0
	NeRF	235	38.0	55.0	49.0

From the provided results, it can be seen that for the metallic and highly reflective object (Industrial dataset), NeRF performs better than photogrammetry, whereas, for the other scenarios, photogrammetry produces more accurate results.

Two other translucent and transparent objects are considered: Bottle_1 and Bottle_2 (Figure 3). Glass objects do not diffusely reflect the incoming light and do not have a texture of their own for photogrammetric 3D reconstruction tasks. Their appearance depends on the object's shape, surrounding background and lighting conditions. Therefore, photogrammetry can easily fail or produce very noisy results in such a situation. On the other hand, NeRF, as declared by Mildenhall et al. [16] can learn to properly generate the geometry associated with transparency due to the view-dependent nature of the NeRF model. For both objects, the photogrammetric- and NeRF-based 3D results are co-registered to the GT data and metrics are computed (Figure 11 and Table 6). Findings prove that NeRF performed better than photogrammetry for transparent objects. For example, the estimated RMSE, STD and MAE for photogrammetry on Bottle_1 are 6.5 mm, 7.1 mm and 7.5 mm, respectively. In contrast, NeRF values were dramatically reduced to 1.3 mm, 1.7 mm, and 2.1 mm, respectively.
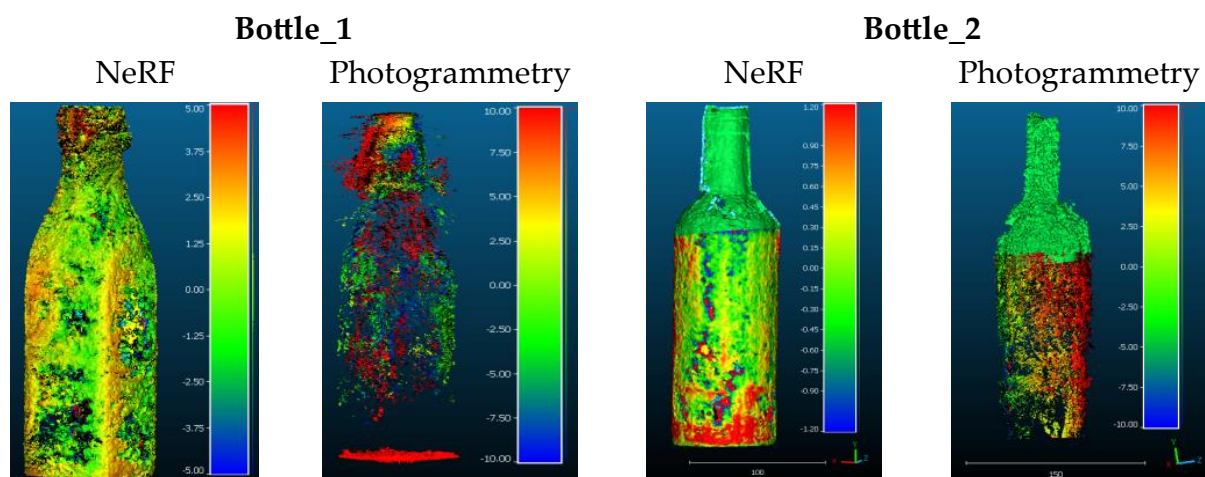

Figure 11. Color-coded cloud-to-cloud comparisons for both Instant-NGP and photogrammetry on the two transparent objects [unit: mm].

Table 6. Statistics of cloud-to-cloud comparisons on transparent objects [unit: mm].

Object	Method	Images	RMSE	MAE	STD
Bottle_1	Pho	360	6.50	7.10	7.50
	NeRF	360	1.30	1.70	2.10
Bottle_2	Pho	360	4.50	5.20	5.10
	NeRF	360	0.63	0.82	1.10

4.7. Accuracy and Completeness

Three different datasets are used to compare photogrammetry and NeRF in terms of accuracy and completeness: Ignatius, Industrial and Bottle_1. For both NeRF (Instant-NGP) and photogrammetry, the two metrics are computed with respect to the available ground truth data. The results, presented in Figure 12, revealed the following insights: (i) for the Ignatius dataset, photogrammetry exhibits higher accuracy and completeness compared to NeRF; (ii) for the Industrial and Bottle_1 datasets, NeRF showcases slightly better results. These findings quantitatively confirm Section 4.6 and that NeRF-based approaches excel when dealing with objects featuring non-collaborative surfaces, particularly those that are transparent or shiny. In contrast, photogrammetry faces challenges in capturing the intricate details of such surfaces, making NeRF a more suitable or complementary choice.
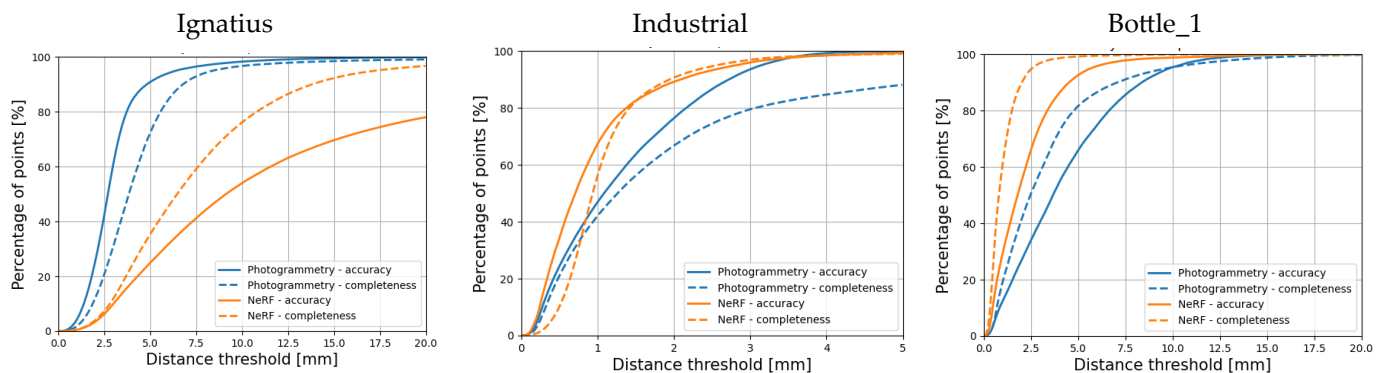
Figure 12. Accuracy and completeness for NeRF and photogrammetry on three different objects.

5. Conclusions

This paper provides a comprehensive analysis of image-based 3D reconstruction using neural radiance field (NeRF) methods. Comparisons with conventional photogrammetry were performed, reporting quantitative and visual results to understand advantages and disadvantages while dealing with multiple types of surfaces and scenes. The study has objectively evaluated the strengths and weaknesses of NeRF-generated 3D data and provided insights into their applicability to different real-life scenarios and applications. The study employed a range of well-textured, texture-less, metallic, translucent, and transparent objects, imaged using different scales and sets of images. The quality of the generated NeRF-based 3D data was evaluated using various evaluation approaches and metrics, including noise level, surface deviation, geometric accuracy, and completeness.

The reported results indicate that NeRF outperforms photogrammetry in scenarios where conventional photogrammetric approaches fail or produce noisy results, such as with texture-less, metallic, highly reflective, and transparent objects. In contrast, photogrammetry still performs better with well-textured and partially textured objects. This is due to the fact that the NeRF-based methods are capable of generating geometry associated with reflectivity and transparency due to the view-dependent nature of the NeRF model.

The study provides valuable insights into the applicability of NeRF for different real-life scenarios, particularly for heritage and industrial scenes, where surfaces can be particularly challenging. More datasets are in preparation and will be shared soon at

https://github.com/3DOM-FBK/NeRFBK [103]. The findings of the study highlight the potentials and limitations of NeRF and photogrammetry, providing a foundation for forthcoming research in this field. Future investigations could explore the combination of NeRF and photogrammetry to improve the quality and efficiency of 3D reconstruction in challenging scenarios.

Author Contributions: Conceptualization, A.K. and F.R.; methodology, A.K. and F.R.; software, A.K., S.R. and Z.Y.; validation, A.K., S.R., G.M. and Z.Y.; formal analysis, A.K. and F.R.; investigation, A.K., G.M., Z.Y. and F.R.; resources, F.R.; data curation, A.K., F.R. and G.M.; writing, original draft preparation, A.K. and F.R.; writing, review and editing, A.K., F.R., G.M., Z.Y., S.R. and R.Q.; visualization, A.K. and Z.Y.; supervision, F.R. and R.Q.; project administration, F.R.; funding acquisition, F.R. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Available upon a reasonable request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Luhmann, T.; Robson, S.; Kyle, S.; Boehm, J. Close-range photogrammetry and 3D imaging. In *Close-Range Photogrammetry and 3D Imaging*; De Gruyter: Berlin, Germany, 2019.
2. Fraser, C.; Brow, D. Industrial photogrammetry: New developments and recent applications. *Photogramm. Rec.* **2006**, *12*, 197–217. [CrossRef]
3. Sansoni, G.; Trebeschi, M.; Docchio, F. State-of-the-art and applications of 3D imaging sensors in industry, cultural heritage, medicine, and criminal investigation. *Sensors* **2009**, *9*, 568–601. [CrossRef] [PubMed]
4. Rodríguez-Martín, M.; Lagüela, S.; González-Aguilera, D.; Rodríguez-Gonzálvez, P. Procedure for quality inspection of welds based on macro-photogrammetric three-dimensional reconstruction. *Opt. Laser Technol.* **2015**, *73*, 54–62. [CrossRef]
5. Karami, A.; Menna, F.; Remondino, F. Combining Photogrammetry and Photometric Stereo to Achieve Precise and Complete 3D Reconstruction. *Sensors* **2022**, *22*, 8172. [CrossRef] [PubMed]
6. Remondino, F. Heritage recording and 3D modeling with photogrammetry and 3D scanning. *Remote Sens.* **2011**, *3*, 1104–1138. [CrossRef]
7. Menna, F.; Nocerino, E.; Remondino, F.; Dellepiane, M.; Callieri, M.; Scopigno, R. 3D digitization of an heritage masterpiece-a critical analysis on quality assessment. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *41*, 675–683. [CrossRef]
8. Stylianidis, E.; Remondino, F. *3D Recording, Documentation and Management of Cultural Heritage*; Whittles Publishing: Dunbeath, UK, 2016; 388p, ISBN 978-184995-168-5.
9. Verhoeven, G.; Wild, B.; Schlegel, J.; Wieser, M.; Pfeifer, N.; Wogrin, S.; Eysn, L.; Carloni, M.; Koschiček-Krombholz, B.; Mola-da-Tebar, A.; et al. Project indigo—Document, disseminate & analyse a graffiti-scape. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2022**, *46*, 513–520.
10. Remondino, F.; El-Hakim, S. Image-based 3D modelling: A review. *Photogramm. Rec.* **2006**, *21*, 269–291. [CrossRef]
11. Remondino, F.; Menna, F.; Koutsoudis, A.; Chamzas, C.; El-Hakim, S. Design and implement a reality-based 3D igitization and modelling project. In Proceedings of the IEEE Conference "Digital Heritage 2013", Marseille, France, 28 October–1 November 2013; Volume 1, pp. 137–144.
12. Wu, B.; Zhou, Y.; Qian, Y.; Gong, M.; Huang, H. Full 3D reconstruction of transparent objects. *arXiv* **2018**, arXiv:1805.03482. [CrossRef]
13. Ahmadabadian, A.H.; Karami, A.; Yazdan, R. An automatic 3D reconstruction system for texture-less objects. *Robot. Auton. Syst.* **2019**, *117*, 29–39. [CrossRef]
14. Santoši, Ž.; Budak, I.; Stojaković, V.; Šokac, M.; Vukelić, Đ. Evaluation of synthetically generated patterns for image-based 3D reconstruction of texture-less objects. *Measurement* **2019**, *147*, 106883. [CrossRef]
15. Karami, A.; Battisti, R.; Menna, F.; Remondino, F. 3D digitization of transparent and glass surfaces: State of the art and analysis of some methods. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2022**, *43*, 695–702. [CrossRef]
16. 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]
17. Tewari, A.; Fried, O.; Thies, J.; Sitzmann, V.; Lombardi, S.; Sunkavalli, K.; Martin-Brualla, R.; Simon, T.; Saragih, J.; Niessner, M.; et al. State of the Art on Neural Rendering. *Comput. Graph. Forum* **2020**, *39*, 701–727. [CrossRef]
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. Wang, X.; Wang, C.; Liu, B.; Zhou, X.; Zhang, L.; Zheng, J.; Bai, X. Multi-view stereo in the deep learning era: A comprehensive review. *Displays* **2021**, *70*, 102102. [CrossRef]
20. 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]
21. Li, Z.; Müller, T.; Evans, A.; Taylor, R.H.; Unberath, M.; Liu, M.Y.; Lin, C.H. Neuralangelo: High-Fidelity Neural Surface Recon-struction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023; pp. 8456–8465.
22. Tancik, M.; Weber, E.; Ng, E.; Li, R.; Yi, B.; Kerr, J.; Wang, T.; Kristoffersen, A.; Austin, J.; Salahi, K.; et al. Nerfstudio: A Modular Framework for Neural Radiance Field Development. *arXiv* **2023**, arXiv:2302.04264.
23. Lorensen, W.E.; Cline, H.E. Marching cubes: A high resolution 3d surface construction algorithm. *ACM Comput. Graph.* **1987**, *21*, 163–169. [CrossRef]
24. Seitz, S.M.; Dyer, C.R. Photorealistic scene reconstruction by voxel coloring. In Proceedings of the IEEE Computer Society Con-ference on Computer Vision and Pattern Recognition, San Juan, PR, USA, 17–19 June 1997; IEEE: New York, NY, USA, 1997; pp. 1067–1073.
25. Agrawal, M.; Davis, L.S. A probabilistic framework for surface reconstruction from multiple images. In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001), Kauai, HI, USA, 8–14 December 2001; IEEE: New York, NY, USA, 2001; Volume 2, p. II.
26. Seitz, S.M.; Curless, B.; Diebel, J.; Scharstein, D.; Szeliski, R. A comparison and evaluation of multi-view stereo reconstruction algorithms. In Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), New York, NY, USA, 17–22 June 2006; IEEE: New York, NY, USA, 2006; Volume 1, pp. 519–528.
27. Bleyer, M.; Rhemann, C.; Rother, C. Patchmatch stereo-stereo matching with slanted support windows. In Proceedings of the BMVC 2011, Dundee, UK, 29 August–2 September 2011; Volume 11, pp. 1–11.
28. 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; Springer International Publishing: Berlin/Heidelberg, Germany, 2016; pp. 501–518.
29. Paschalidou, D.; Ulusoy, O.; Schmitt, C.; Van Gool, L.; Geiger, A. Raynet: Learning volumetric 3d reconstruction with ray po-tentials. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 3897–3906.
30. Morelli, L.; Bellavia, F.; Menna, F.; Remondino, F. PHOTOGRAMMETRY NOW AND THEN–FROM HAND-CRAFTED TO DEEP-LEARNING TIE POINTS–. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2022**, *48*, 163–170. [CrossRef]
31. Ulusoy, A.O.; Geiger, A.; Black, M.J. Towards probabilistic volumetric reconstruction using ray potentials. In Proceedings of the 2015 International Conference on 3D Vision, Lyon, France, 19–22 October 2015; IEEE: New York, NY, USA, 2015; pp. 10–18.
32. Tulsiani, S.; Zhou, T.; Efros, A.A.; Malik, J. Multi-view supervision for single-view reconstruction via differentiable ray con-sistency. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2626–2634.
33. Zagoruyko, S.; Komodakis, N. Learning to compare image patches via convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 4353–4361.
34. Luo, W.; Schwing, A.G.; Urtasun, R. Efficient deep learning for stereo matching. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 5695–5703.
35. Ummenhofer, B.; Zhou, H.; Uhrig, J.; Mayer, N.; Ilg, E.; Dosovitskiy, A.; Brox, T. Demon: Depth and motion network for learning monocular stereo. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 5038–5047.
36. Leroy, V.; Franco, J.S.; Boyer, E. Shape reconstruction using volume sweeping and learned photoconsistency. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 781–796.
37. Riegler, G.; Ulusoy, A.O.; Bischof, H.; Geiger, A. Octnetfusion: Learning depth fusion from data. In Proceedings of the 2017 International Conference on 3D Vision (3DV), Qingdao, China, 10–12 October 2017; IEEE: New York, NY, USA, 2017; pp. 57–66.
38. Donne, S.; Geiger, A. Learning non-volumetric depth fusion using successive reprojections. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 7634–7643.
39. Huang, P.H.; Matzen, K.; Kopf, J.; Ahuja, N.; Huang, J.B. Deepmvs: Learning multi-view stereopsis. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 2821–2830.
40. Yao, Y.; Luo, Z.; Li, S.; Shen, T.; Fang, T.; Quan, L. Recurrent mvsnet for high-resolution multi-view stereo depth inference. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 5525–5534.
41. Yu, Z.; Gao, S. Fast-mvsnet: Sparse-to-dense multi-view stereo with learned propagation and gauss-newton refinement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 1949–1958.
42. Menna, F.; Nocerino, E.; Morabito, D.; Farella, E.M.; Perini, M.; Remondino, F. An open source low-cost automatic system for image-based 3D digitization. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2017**, *42*, 155. [CrossRef]

43. Mousavi, V.; Khosravi, M.; Ahmadi, M.; Noori, N.; Haghshenas, S.; Hosseininaveh, A.; Varshosaz, M. The performance evalu-ation of multi-image 3D reconstruction software with different sensors. *Measurement* **2018**, *120*, 1–10. [CrossRef]

44. Hafeez, J.; Lee, J.; Kwon, S.; Ha, S.; Hur, G.; Lee, S. Evaluating feature extraction methods with synthetic noise patterns for image-based modelling of texture-less objects. *Remote Sens.* **2020**, *12*, 3886. [CrossRef]

45. Nicolae, C.; Nocerino, E.; Menna, F.; Remondino, F. Photogrammetry applied to problematic artefacts. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2014**, *40*, 451. [CrossRef]

46. Wallis, R. An approach to the space variant restoration and enhancement of images. In Proceedings of the Symposium on Current Mathematical Problems in Image Science, Naval Postgraduate School, Monterey, CA, USA, 10-12 November 1976.

47. Gaiani, M.; Remondino, F.; Apollonio, F.; Ballabeni, A. An advanced pre-processing pipeline to improve automated photo-grammetric reconstructions of architectural scenes. *Remote Sens.* **2016**, *8*, 178. [CrossRef]

48. Karami, A.; Menna, F.; Remondino, F. Investigating 3D reconstruction of non-collaborative surfaces through photogrammetry and photometric stereo. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2021**, *43*, 519–526. [CrossRef]

49. Karami, A.; Menna, F.; Remondino, F.; Varshosaz, M. Exploiting light directionality for image-based 3d reconstruction of non-collaborative surfaces. *Photogramm. Rec.* **2022**, *37*, 111–138. [CrossRef]

50. Park, J.; Sinha, S.N.; Matsushita, Y.; Tai, Y.W.; Kweon, I.S. Robust igitizat photometric stereo using planar mesh parameterization. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 1591–1604. [CrossRef]

51. Logothetis, F.; Mecca, R.; Cipolla, R. A differential volumetric approach to multi-view photometric stereo. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 1052–1061.

52. Li, M.; Zhou, Z.; Wu, Z.; Shi, B.; Diao, C.; Tan, P. Multi-view photometric stereo: A robust solution and benchmark dataset for spatially varying isotropic materials. *IEEE Trans. Image Process.* **2020**, *29*, 4159–4173. [CrossRef]

53. Karami, A.; Varshosaz, M.; Menna, F.; Remondino, F.; Luhmann, T. FFT-based Filtering Approach to Fuse Photogrammetry and Photometric Stereo 3D Data. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2023**, *14*, 363–370. [CrossRef]

54. Woodham, R.J. Photometric method for determining surface orientation from multiple images. *Opt. Eng.* **2021**, *19*, 191139. [CrossRef]

55. Antensteiner, D.; Štolc, S.; Pock, T. A review of depth and normal fusion algorithms. *Sensors* **2018**, *18*, 431. [CrossRef] [PubMed]

56. Chen, G.; Han, K.; Wong, K.K. PS-FCN: A Flexible Learning Framework for Photometric Stereo. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.

57. Ju, Y.; Lam, K.M.; Chen, Y.; Qi, L.; Dong, J. Pay Attention to Devils: A Photometric Stereo Network for Better Details. In Proceedings of the 29th International Conference on International Joint Conferences on Artificial Intelligence, Yokohama, Japan, 11–17 July 2020.

58. Logothetis, F.; Budvytis, I.; Mecca, R.; Cipolla, R. PX-NET: Simple and Efficient Pixel-Wise Training of Photometric Stereo Networks. *arXiv* **2021**, arXiv:2008.04933v3.

59. Tewari, A.; Thies, J.; Mildenhall, B.; Srinivasan, P.; Tretschk, E.; Yifan, W.; Lassner, C.; Sitzmann, V.; Martin-Brualla, R.; Lombardi, S.; et al. Advances in Neural Rendering. *Comput. Graph. Forum* **2022**, *41*, 703–735. [CrossRef]

60. Niemeyer, M.; Barron, J.T.; Mildenhall, B.; Sajjadi, M.S.; Geiger, A.; Radwan, N. Regnerf: Regularizing neural radiance fields for view synthesis from sparse inputs. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 5480–5490.

61. Deng, K.; Liu, A.; Zhu, J.Y.; Ramanan, D. Depth-supervised nerf: Fewer views and faster training for free. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 12882–12891.

62. Guo, Y.C.; Kang, D.; Bao, L.; He, Y.; Zhang, S.H. NeRFren: Neural radiance fields with reflections. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 18409–18418.

63. Zhang, X.; Srinivasan, P.P.; Deng, B.; Debevec, P.; Freeman, W.T.; Barron, J.T. Nerfactor: Neural factorization of shape and reflectance under an unknown illumination. *ACM Trans. Graph. ToG* **2021**, *40*, 1–18. [CrossRef]

64. 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, New Orleans, LA, USA, 18–24 June 2022; pp. 5470–5479.

65. Zhan, H.; Zheng, J.; Xu, Y.; Reid, I.; Rezatofighi, H. ActiveRMAP: Radiance Field for Active Mapping and Planning. *arXiv* **2022**, arXiv:2211.12656.

66. Gao, K.; Gao, Y.; He, H.; Lu, D.; Xu, L.; Li, J. NeRF: Neural radiance field in 3d vision, a comprehensive review. *arXiv* **2022**, arXiv:2210.00379.

67. Kolodiazhna, O.; Savin, V.; Uss, M.; Kussul, N. 3D Scene Reconstruction with Neural Radiance Fields (NeRF) Considering Dy-namic Illumination Conditions. In Proceedings of the International Conference on Applied Innovation in IT, Manchester, UK, 13–15 April 2023; Anhalt University of Applied Sciences: Köthen, Germany, 2023; Volume 11, pp. 233–238.

68. Reiser, C.; Szeliski, R.; Verbin, D.; Srinivasan, P.P.; Mildenhall, B.; Geiger, A.; Barron, J.T.; Hedman, P. Merf: Memory-efficient radiance fields for real-time view synthesis in unbounded scenes. *arXiv* **2023**, arXiv:2302.12249.

69. Tancik, M.; Srinivasan, P.; Mildenhall, B.; Fridovich-Keil, S.; Raghavan, N.; Singhal, U.; Ramamoorthi, R.; Barron, J.; Ng, R. Fourier features let networks learn high frequency functions in low dimensional domains. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 7537–7547.

70. Sitzmann, V.; Martel, J.; Bergman, A.; Lindell, D.; Wetzstein, G. Implicit neural representations with periodic activation functions. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 7462–7473.

71. Sun, C.; Sun, M.; Chen, H.T. Direct voxel grid optimization: Super-fast convergence for radiance fields reconstruction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 5459–5469.

72. Chen, A.; Xu, Z.; Geiger, A.; Yu, J.; Su, H. Tensorf: Tensorial radiance fields. In Proceedings of the Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, 23–27 October 2022; Proceedings, Part XXXII. Springer Nature: Cham, Switzerland, 2022; pp. 333–350.

73. Rebain, D.; Jiang, W.; Yazdani, S.; Li, K.; Yi, K.M.; Tagliasacchi, A. Derf: Decomposed radiance fields. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19-25 June 2021; pp. 14153–14161.

74. Takikawa, T.; Evans, A.; Tremblay, J.; Müller, T.; McGuire, M.; Jacobson, A.; Fidler, S. Variable bitrate neural fields. In Proceedings of the ACM SIGGRAPH 2022 Conference Proceedings, Vancouver, BC, Canada, 7–11 August 2022; pp. 1–9.

75. Boss, M.; Braun, R.; Jampani, V.; Barron, J.T.; Liu, C.; Lensch, H. Nerd: Neural reflectance decomposition from image collections. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Nashville, TN, USA, 20–25 June 2021; pp. 12684–12694.

76. Srinivasan, P.P.; Deng, B.; Zhang, X.; Tancik, M.; Mildenhall, B.; Barron, J.T. Nerv: Neural reflectance and visibility fields for relighting and view synthesis. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Montreal, BC, Canada, 11–17 October 2021; pp. 7495–7504.

77. Verbin, D.; Hedman, P.; Mildenhall, B.; Zickler, T.; Barron, J.T.; Srinivasan, P. Ref-NeRF: Structured view-dependent ap-pearance for neural radiance fields. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recog-nition (CVPR), New Orleans, LA, USA, 19–20 June 2022; IEEE: New York, NY, USA, 2022; pp. 5481–5490.

78. Zhu, Z.; Peng, S.; Larsson, V.; Xu, W.; Bao, H.; Cui, Z.; Oswald, M.R.; Pollefeys, M. Nice-slam: Neural implicit scalable encoding for slam. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–20 June 2022.

79. 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.

80. Yang, W.; Chen, G.; Chen, C.; Chen, Z.; Wong, K.Y.K. Ps-NeRF: Neural inverse rendering for multi-view photometric stereo. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, 23–27 October 2022*; Proceedings, Part I; Springer Nature: Cham, Switzerland, 2022; pp. 266–284.

81. Park, K.; Sinha, U.; Barron, J.T.; Bouaziz, S.; Goldman, D.B.; Seitz, S.M.; Martin-Brualla, R. Nerfies: Deformable neural radiance fields. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 5865–5874.

82. Pumarola, A.; Corona, E.; Pons-Moll, G.; Moreno-Noguer, F. D-NeRF: Neural Radiance Fields for Dynamic Scenes. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Montreal, BC, Canada, 11–17 October 2021.

83. Wu, T.; Zhong, F.; Tagliasacchi, A.; Cole, F.; Oztireli, C. D2 NeRF: Self-Supervised Decoupling of Dynamic and Static Objects from a Monocular Video. *arXiv* **2022**, arXiv:2205.15838.

84. 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, 18–22 June 2023; pp. 8285–8295.

85. Cao, A.; Johnson, J. Hexplane: A fast representation for dynamic scenes. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023; pp. 130–141.

86. Jain, A.; Tancik, M.; Abbeel, P. Putting NeRF on a diet: Semantically consistent few-shot view synthesis. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Nashville, TN, USA, 20–25 June 2021; pp. 5885–5894.

87. 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.

88. Zhang, K.; Riegler, G.; Snavely, N.; Koltun, V. NeRF++: Analyzing and improving neural radiance fields. *arXiv* **2020**, arXiv:2010.07492.

89. Guo, H.; Peng, S.; Lin, H.; Wang, Q.; Zhang, G.; Bao, H.; Zhou, X. Neural 3D scene reconstruction with the igitizat-world as-sumption. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 5511–5520.

90. Yu, Z.; Peng, S.; Niemeyer, M.; Sattler, T.; Geiger, A. MonoSDF: Exploring Monocular Geometric Cues for Neural Implicit Surface Reconstruction. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 25018–25032.

91. Bian, W.; Wang, Z.; Li, K.; Bian, J.W.; Prisacariu, V.A. Nope-nerf: Optimising neural radiance field with no pose prior. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023; pp. 4160–4169.

92. Rakotosaona, M.J.; Manhardt, F.; Arroyo, D.M.; Niemeyer, M.; Kundu, A.; Tombari, F. NeRFMeshing: Distilling Neural Radiance Fields into Geometrically-Accurate 3D Meshes. *arXiv* **2023**, arXiv:2303.09431.

93. Elsner, T.; Czech, V.; Berger, J.; Selman, Z.; Lim, I.; Kobbelt, L. Adaptive Voronoi NeRFs. *arXiv* **2023**, arXiv:2303.16001.

94. Kulhanek, J.; Sattler, T. Tetra-NeRF: Representing Neural Radiance Fields Using Tetrahedra. *arXiv* **2023**, arXiv:2304.09987.

95. Yu, A.; Ye, V.; Tancik, M.; Kanazawa, A. pixelNeRF—Neural Radiance Fields from One or Few Images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Montreal, BC, Canada, 11–17 October 2021.

96. Croce, V.; Caroti, G.; De Luca, L.; Piemonte, A.; Véron, P. Neural Radiance Fields (NeRF): Review and potential applications to Digital Cultural Heritage. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2023**, *48*, 453–460.

97. Mazzacca, G.; Karami, A.; Rigon, S.; Farella, E.M.; Trybala, P.; Remondino, F. NeRF for heritage 3D reconstruction. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2023**, *48*, 1051–1058. [CrossRef]

98. Yu, Z.; Chen, A.; Antic, B.; Peng, S.P.; Bhattacharyya, A.; Niemeyer, M.; Tang, S.; Sattler, T.; Geiger, A. SDFStudio: A Unified Framework for Surface Reconstruction. Available online: https://github.com/autonomousvision/sdfstudio (accessed on 1 July 2023).

99. Besl, P.J.; McKay, N.D. A method for registration of 3D shapes. *IEEE Trans. Pattern Anal. Mach. Intell.* **1992**, *14*, 239–256. [CrossRef]

100. Mohammadi, M.; Rashidi, M.; Mousavi, V.; Karami, A.; Yu, Y.; Samali, B. Quality evaluation of digital twins generated based on UAV photogrammetry and TLS: Bridge case study. *Remote Sens.* **2021**, *13*, 3499. [CrossRef]

101. Knapitsch, A.; Park, J.; Zhou, Q.Y.; Koltun, V. Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM Trans. Graph. ToG* **2017**, *36*, 1–13. [CrossRef]

102. Nocerino, E.; Stathopoulou, E.K.; Rigon, S.; Remondino, F. Surface reconstruction assessment in photogrammetric applications. *Sensors* **2020**, *20*, 5863. [CrossRef] [PubMed]

103. NeRFBK Dataset. Available online: https://github.com/3DOM-FBK/NeRFBK/ (accessed on 1 July 2023).

104. Harrison, R.L. Introduction to monte carlo simulation. *AIP Conf. Proc.* **2021**, *1204*, 17–21.