

Using NeRFs for UAV-based 3D reconstruction of complex scenes: A comparison to MVS

Frederik Schulte¹, Perpetual Akwensi¹, Lukas Winiwarter¹

¹ Unit of Geometry and Surveying, University of Innsbruck, Austria - (frederik.schulte, perpetual.akwensi, lukas.winiwarter)@uibk.ac.at

Keywords: UAV photogrammetry, NeRF, 3DGS, Multi-View Stereo, Point cloud evaluation.

Abstract

High-resolution 3D documentation of cultural heritage sites is essential for their preservation. While terrestrial laser scanning (TLS) remains the gold standard, it is often cost-intensive compared to photogrammetry. This study evaluates three image-based reconstruction techniques, Multi-View Stereo (MVS), Neural Radiance Fields (NeRF) and 3D Gaussian Splatting (3DGS), by applying them to a complex scene featuring a chapel and its surrounding vegetation, sensed from an uncrewed aerial vehicle (UAV). A hybrid TLS/MVS model provides a high-accuracy reference. Using identical interior and exterior camera parameters of the 105 UAV-acquired images, we generate dense point clouds with all methods and assess geometric accuracy and completeness using the M3C2 algorithm. Results show that MVS achieves superior accuracy (standard deviation of all M3C2 distances: MVS = 0.11 m, NeRF = 0.15 m), whereas NeRF attains up to 20% higher completeness, particularly in low-texture and vegetation-occluded regions. The 3DGS point cloud was deemed too sparse and was therefore not used for further analysis. The study highlights the potential of NeRFs to recover partially occluded or sparsely textured geometries that are challenging for MVS and suggests a complementary use of both approaches for cost-efficient documentation of cultural heritage.

1. Introduction

The precise documentation of cultural heritage sites represents an important task, essential for preserving cultural memory. Over the past decade, numerous historical sites have been three-dimensionally recorded and archived using geodetic methods (Aicardi et al., 2018). Especially in alpine regions, many of these sites are increasingly threatened due to the effects of climate change, as gravitational mass movements such as landslides, slope failures, debris flows, and rockfalls in the Alps are becoming more frequent and severe (Huggel et al., 2012). These hazards threaten such sites both directly, as the ground on which the structures stand may shift and cause structural damage, and indirectly, through the risk of being buried by unstable terrain located above them. To effectively document and monitor these areas and structures, the acquired 3D surface data must meet two key criteria: Firstly, a high-resolution, geometrically accurate point cloud with quantified uncertainty to detect small deformations (Winiwarter et al., 2021), and secondly, high completeness to guarantee that the entire structure is captured for archival documentation. The reconstructed 3D model can then also be used for modern structural analysis with finite element methods (Gonizzi Barsanti et al., 2022) or visualization frameworks such as extended reality applications (Teruggi et al., 2021). One of the challenges is to combine the disciplines of geodetic monitoring and documentation of cultural heritage within a unified methodological framework.

In most cases, terrestrial laser scanning (TLS) or mobile laser scanning (MLS) are the technologies of choice to acquire such data, as they deliver reliable 3D surface models (Irwin et al., 2025). However, the instruments are expensive, motivating the development of more cost-effective approaches. In recent years, unmanned aerial vehicles (UAVs) with cameras have become significantly more affordable and thus represent a viable alternative for monitoring alpine infrastructures and terrain. To bridge the quality gap between laser scanning and photogrammetry and to enable faster and more cost-efficient data acquisition, a promising approach has recently emerged. Instead

of relying solely on classical photogrammetric algorithms, 3D scenes can now be reconstructed using radiance-field-based methods. This approach offers the potential to optimize the reconstruction difficulties in regions with poor structure or movement of objects between images.

Radiance field techniques have rapidly risen to prominence in computer vision research and are now being more and more adopted in engineering disciplines. Classical photogrammetric reconstruction algorithms, most notably Multi-View Stereo (MVS) (Furukawa and Ponce, 2007), suffer from pronounced limitations when applied to alpine environment: Low-texture surfaces and sparsely discretized vegetation impede reliable 3D reconstruction. In contrast, point clouds generated from Neural Radiance Fields (NeRFs) (Mildenhall et al., 2021) often exhibit higher completeness (Schulte et al., 2025). The advantage of NeRFs lies in their implicit geometry representation. This approach effectively captures fine-scale, sub-pixel geometries and adeptly models dynamic elements (like water surfaces or moving vegetation) and texture-less regions present in the images the NeRF is trained on (Remondino et al., 2023). NeRFs also outperform MVS in texture-poor scenarios and have a higher robustness when vegetation occlusion appears in the images (Petrovska and Jutzi, 2025).

This study therefore investigates the performance of radiance-field-based point clouds, like NeRFs and 3D Gaussian Splatting (3DGS) (Kerbl et al., 2023), compared to classical MVS point clouds. For this purpose, a – from a photogrammetric perspective complex – scene was selected in which the accuracy and completeness of both radiance-field-based and classical approaches are evaluated against a TLS reference point cloud. Although the scene is not complex in terms of its materials, some areas are difficult to see clearly in multiple images, and there are moving vegetation elements present. Therefore, it is likely that MVS will have difficulties in reconstructing these areas.

The focus of this study lies on geometry extraction in a realistic surveying application, where the exported 3D geometry,

rather than 2D view synthesis is the primary object of investigation. In the recent literature, several studies have examined the use of radiance fields to reconstruct geometry of objects that are related to cultural heritage. However, works such as Yan et al. (2025) and Billi et al. (2025) employ NeRFs for the reconstruction of small objects, which is not directly comparable to the acquisition of an entire building. Other studies, including Xiang et al. (2025) and Condorelli et al. (2021), reconstruct comparable scenes using radiance fields, but without validating them against a reference measurement, which allows only limited conclusions regarding reconstruction quality. Clini et al. (2024), in contrast, conducted a more comparable study in which a church was surveyed using a DJI Mavic Mini UAV. The evaluation was performed using MVS, NeRF, and 3DGS and was geometrically assessed against a TLS reference. However, only a single facade was considered, and no occlusions were present. Our article focuses in particular on a challenging, realistic scenario, such as those often encountered with many properties in rural areas where the buildings are occluded by vegetation or other buildings. Moreover in Clini et al. (2024), completeness of the reconstruction was not assessed in this research, although previous studies suggest that NeRFs may offer an advantage by producing more complete, albeit less accurate, scene reconstructions (Schulte et al., 2026).

Another possibility to test radiance-fields-based workflows would have been the use of existing benchmarks such as *Tanks and Temples* (Knapitsch et al., 2017) or *NeRFBook* (Yan et al., 2023). However, these are computer-vision-oriented scenes that do not meet surveying standards. On the one hand, they typically rely on smartphone cameras, whose geometric imaging quality does not match that of sensors commonly used in photogrammetry, which limits their transferability to surveying applications. On the other hand, the data is not georeferenced.

Therefore, this contribution presents a newly acquired dataset and a surveying-grade georeferencing pipeline to process high-resolution georeferenced data. The geometric evaluation is carried out on an entire chapel including the surrounding vegetation, using a high-precision TLS measurement as reference.

2. Methodology

In this section, the fundamentals of the employed reconstruction techniques are presented. Section 2.1 gives an overview of the Structure-from-Motion (SfM) workflow. Following this, the dense point cloud generation using MVS is described in Section 2.2, the implicit radiance-field-based approach using NeRFs in Section 2.3, and the explicit point-based rendering via 3DGS in Section 2.4.

2.1 Image orientation and alignment

SfM is employed to determine the internal and external orientation parameters of a camera necessary for 3D reconstruction (Schönberger and Frahm, 2016). Distinctive image features are extracted, for example with the Scale Invariant Feature Transform (SIFT) operator (Lowe, 2004), and matched in overlapping image pairs using an outlier-robust method such as the Random Sample Consensus (RANSAC) algorithm (Fischler and Bolles, 1981). From the matched key points, termed tie points, relative camera orientations are estimated, followed by a global bundle adjustment. The bundle adjustment simultaneously minimizes a) the reprojection error, by optimizing the exterior orientations, defined by the camera positions (X, Y, Z)

and rotation angles (ω, ϕ, κ), and b) the internal camera parameters (Triggs et al., 2000). These internal parameters describe both the location of the projection center in the image space and the deviations of the actual recording geometry from the ideal central projection (Luhmann, 2010). Optimization of the 2D tie point observations within the image space during bundle adjustment concurrently facilitates the triangulation of a sparse 3D tie point cloud.

2.2 Dense point cloud generation with MVS

To generate a dense point cloud from the sparse SfM point cloud and the estimated camera orientations, classical MVS algorithms are applied (Schönberger et al., 2016). A typical MVS pipeline constrains the correspondence search based on epipolar geometry, restricting potential matches for a given pixel along its corresponding epipolar line. Dense correspondences are established by evaluating the consistency of local image parts across overlapping views. This matching process yields a dense depth map for each photo. Subsequently, every depth map is projected into a 3D coordinate system to generate a per-view point cloud. These individual point clouds are then fused and filtered to produce a single, globally consistent dense point cloud (Furukawa and Hernández, 2015). On well-textured surfaces and under stable lighting, MVS can be applied and parallelized to very large datasets efficiently (Galliani et al., 2015), producing results that are accurate, comparable to TLS (Wilkinson et al., 2016). In contrast, quality of the reconstruction degrades significantly on textureless, reflective, or refractive surfaces, in poor lighting, or in complex scenes with low image overlap.

2.3 Neural Radiance Fields (NeRFs)

In contrast to the 3D point cloud derived using MVS, NeRFs represent 3D scenes implicitly by applying deep learning to overlapping images. Like MVS, the internal and external orientation parameters of the cameras are prerequisites. Then, a neural network that constructs a radiance field, thereby enabling the construction of novel views from arbitrary perspectives (view synthesis), is trained. In contrast to classical geodetic methods that explicitly store 3D coordinates of the points, NeRFs encode scene geometry and appearance implicitly within the learned network weights.

In the original NeRF formulation (Mildenhall et al., 2021), the machine learning input is a 5D vector consisting of a 3D spatial position (x, y, z) and a 2D viewing direction (θ, ϕ). Points are sampled along each camera ray, and for each sample, the model outputs a color (R, G, B) and a volume density σ . During training, synthetic pixels are rendered by integrating these values along the rays according to the standard volume rendering principles (Yariv et al., 2021), where σ defines the termination probability of the ray at a given location. The parameters of the neural network are optimized by minimizing the rendering loss between synthetic and observed pixels, exploiting the differentiability of the volume-rendering equation to refine the multilayer perceptron until the scene is accurately reconstructed (Niemeyer et al., 2020).

To obtain a 3D point cloud comparable to MVS output, this geometry is extracted from the trained radiance field in an additional step. During point cloud extraction, rays from the training images are traced through the radiance field until the predicted cumulative volume density exceeds a predefined threshold, at which point the corresponding RGB value is recorded as a surface sample (Oechsle et al., 2021).

2.4 3D Gaussian Splatting (3DGS)

To overcome the computational limitations of implicit neural rendering methods such as NeRFs while maintaining high visual quality, 3DGS was introduced as an explicit scene representation (Kerbl et al., 2023). In contrast to NeRFs, which represent a scene as a continuous volumetric function encoded in neural network parameters, 3DGS models the environment using a set of anisotropic 3D Gaussian primitives.

Training typically also relies on known internal and external camera parameter as well as an initial sparse point cloud, all of which are commonly obtained through an SfM pipeline (Section 2.1). Each Gaussian is parameterized by its 3D center position (μ), a covariance matrix (Σ) describing its scale and orientation, an opacity value (α) and Spherical Harmonics (SH) coefficients to model view-dependent color. A key difference to NeRF lies in the rendering process. Instead of computationally expensive volumetric ray marching through a neural network, 3DGS employs a fast, visibility-aware, differentiable tile-based rasterizer. This approach projects ("splats") the 3D Gaussians onto the 2D image plane and accumulates their contributions via α -blending (Kerbl et al., 2023). During optimization, an adaptive density control mechanism refines the representation by cloning, splitting, and pruning Gaussians based on their contribution to the reconstruction quality. This explicit formulation enables faster training times and real-time rendering at high frame rates, depending on scene complexity and hardware configuration.

Furthermore, 3DGS provides a flexible representation that can capture fine geometric details and view-dependent appearance without requiring an explicit surface mesh, making it particularly suitable for complex scenes with thin structures or non-Lambertian effects. While geometric information can be approximated by extracting the Gaussian centers (μ) as a point-based representation, this does not correspond to a classical dense point cloud, as the Gaussians represent spatially extended volumetric primitives with anisotropic support and view-dependent radiometric properties rather than discrete surface samples.

3. Data acquisition

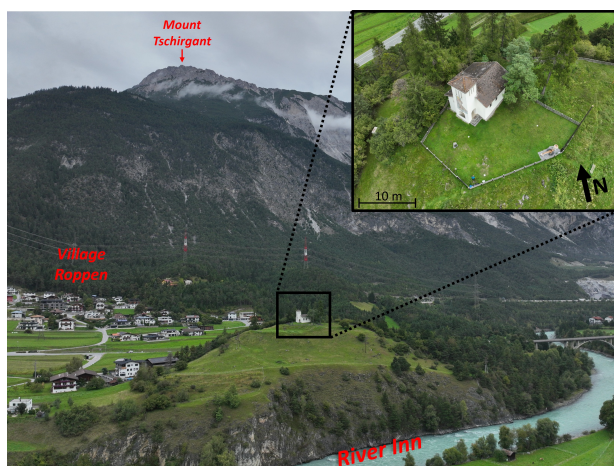


Figure 1. Overview of the study site.

To illustrate the differences between MVS and NeRF reconstructions, we select the *Bruder Klaus Kapelle* in Roppen, Aus-

tria (WGS84: 47.2213°N, 10.8168°E), as our object of interest (Figure 1). The chapel, inaugurated in 1962, is of manageable dimensions (10 m × 12 m × 10 m) yet offers a variety of acquisition scenarios. To the southwest lies an open meadow that provides an unobstructed view for oblique image acquisition, allowing a precise comparison of fine structural reconstruction. In contrast, the northeastern side is covered by dense vegetation, which significantly complicates the reconstruction of the walls using image-based methods. That side in particular enables a meaningful evaluation of point cloud completeness under challenging conditions.

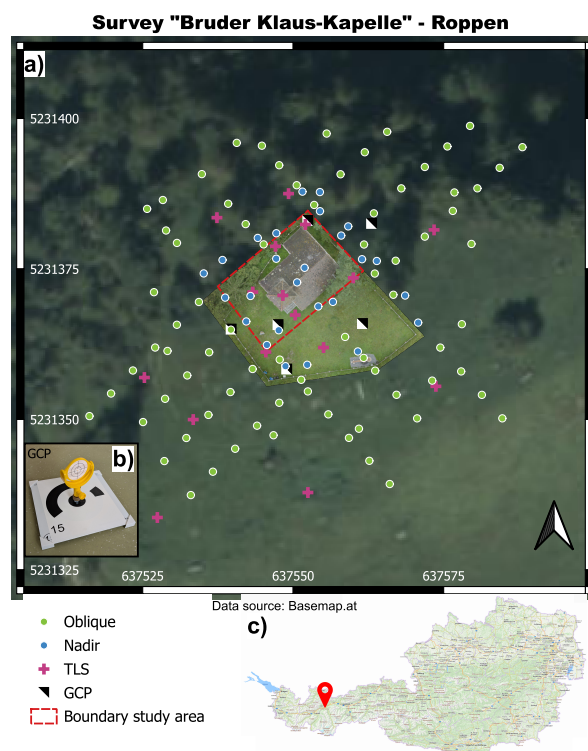


Figure 2. a) Overview of the surveyed area. The positions of the images, separated into oblique and nadir, are displayed together with the TLS positions and GCPs. The study area is also marked. The grid coordinates refer to WGS84/UTM32. b) Close-up of one of the GCPs. c) Position of the study area shown on a map of Austria.

The UAV platform used was a DJI Mavic 3M. Its primary RGB camera is equipped with a 20 MP 4/3 CMOS sensor and an equivalent focal length of 24 mm (DJI Enterprise, 2025). The overlap was 80% in the longitudinal direction and 60% in the transverse direction. Images were captured in RAW format and the flight speed was set to 1 m/s, allowing the use of the highest possible aperture (mostly F/8) and the lowest ISO under the prevailing lighting conditions to achieve high image quality (CANON INC., 2006). A short exposure time of 1/1000 s was chosen, rendering motion blur negligible. To simulate a realistic measurement campaign, the *Smart Oblique* mode with terrain following was employed at a flight height of 25 m above ground level (a.g.l.) (DJI Enterprise, 2021). In this mode, oblique images with an elevation angle of 45° are acquired in addition to nadir images, enabling the reconstruction of facades and lower vegetation. Additionally, a nadir mission and a circular flight focusing on the chapel were conducted at 35 m a.g.l..

For reference measurements, a combined approach of TLS

and photogrammetric evaluation was employed. The area was scanned with a Riegl VZ-600i (Riegl, 2025b) from 16 scan positions. To register the images and the laser scan to a common coordinate system, six specially developed ground control points (GCPs) were distributed throughout the study area. Additionally the RTK positions from UAV and TLS were used. These consisted of 12-bit coded markers that can be automatically detected in the images, incorporating retro-reflective targets that were fine-scanned with the TLS. Since all GCPs were placed horizontally and the vertical offset between the marker center and the bireflex center is known, an accurate registration was achieved. A photograph of a sample GCP is shown in Figure 2b.

4. Data processing

In this section, we adopt the workflow introduced by Schulte et al. (2025) (see Figure 3). Section 4.1 details the processing of the reference point cloud, Section 4.2 describes the estimation of camera poses via SfM, Section 4.3 details the construction of a dense MVS point cloud, Section 4.4 explains the generation of a NeRF-derived point cloud and Section 4.5 describes the process using 3DGS.

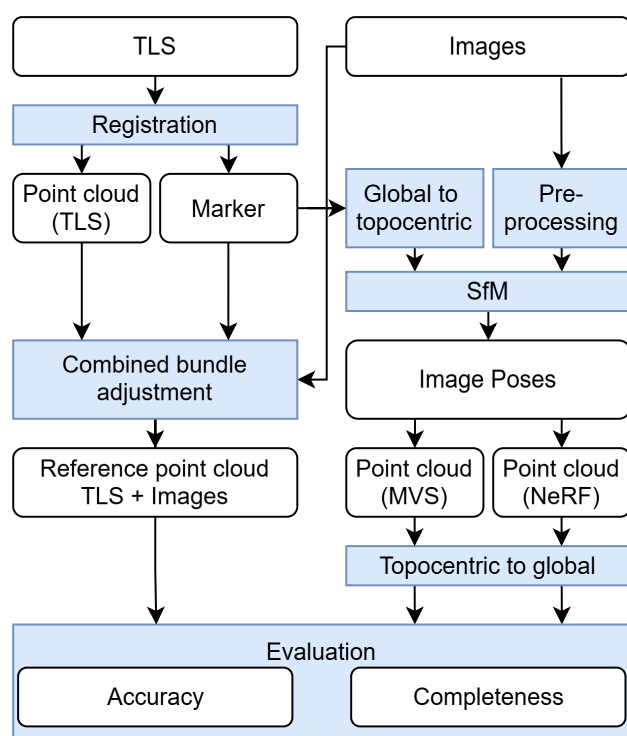


Figure 3. The workflow describes the generation of the reference point cloud from TLS and images in a joint bundle adjustment on the left side. The fused reference point cloud serves as the basis for the evaluation of image-based reconstruction with MVS and NeRF on the right side.

4.1 Reference point cloud generation

To generate the reference point cloud, the TLS data was processed first. TLS processing was performed with the scanner manufacturer’s software RiSCAN PRO (V2.21) (Riegl, 2025a). The scanner already successfully registered all scans relative to each other in the field (Ullrich and Fürst, 2017). The preregistered scans were imported with all returns (multiple

echoes) to best represent the surrounding vegetation. By examining and filtering the input signal according to its pulse shape (max deviation = 10), we prevented incorrect reconstructions due to heavily distorted waveforms. Subsequently, the coordinates of the scanned bireflex targets were calculated using the polar elements saved during the fine scan of these targets. Because each GCP was scanned two or three times, the local precision of each point could be estimated. The mean local uncertainty was 2 mm with a maximum of 4 mm. This can be mainly attributed to manual reflector rotation between scans.

The scans and GCP coordinates were then processed in a multi-station adjustment (MSA). At its core, this adjustment uses plane-to-plane matching (Rietdorf, 2005) to register the scans relative to each other at sub-millimeter precision and subsequently georeferences the registered dataset using the RTK-positions of the individual stations in the block adjustment. Only the vertical offset between the bireflex target center and the coded-marker center remained to be applied. Because all GCPs were nearly leveled, this offset is limited to the z-axis and could therefore be corrected easily. As a result, precise GCP coordinates were obtained for the subsequent analyses.

Because topography prevented complete TLS coverage of all roof surfaces, a combined TLS/photogrammetry workflow employed using Agisoft Metashape (V2.2.2) to obtain a complete and reliable reference point cloud (Agisoft LLC, 2025). The final registered TLS point cloud served as the geometric basis by fixing the scan positions. The relative orientations of all 105 images (oblique + nadir) were then estimated and, together with the TLS scans, optimized in a joint bundle adjustment. A dense point cloud was then generated from the full-resolution images using MVS. Points that are reconstructed from fewer than three depth maps were removed to reduce noise, and the resulting MVS point cloud was cropped to the chapel area, so that only the well-textured roof surfaces were created using photogrammetric methods. Finally, the combined point cloud (TLS+MVS) was cropped to the study area and exported in the global coordinate reference system (CRS), producing a high-accuracy, comprehensive reference point cloud of the chapel and the surrounding vegetation containing approximately 139 million points.

4.2 Camera pose estimation

To compare the geometric quality of the MVS and NeRF methods, both were given identical image datasets, camera poses and intrinsics as input. To ensure optimal results, the RAW images were first radiometrically balanced to prevent overexposure of the bright chapel facade against the darker background, and then exported as lightly compressed JPGs. Because both reconstruction methods require internal and external orientation parameters, these were estimated using an SfM approach. The workflow followed the procedure described by Over et al. (2021), implemented through the Agisoft Metashape plugin created by Mohren and Schulze (2024).

The bundle adjustment was performed iteratively using the GCP coordinates derived from the TLS target fine scans. To efficiently train a NeRF, the markers were transformed into a local topocentric coordinate system, ensuring that the camera poses estimated in the bundle adjustment also use this local system. This approach offers two advantages. First, the NeRF can be trained in a numerically more resource-efficient local coordinate system, and second, using global coordinates during training

led to an unresolved issue in which the exported point cloud became excessively noisy, due to wrong exterior orientation parameters. Using a local CRS additional avoids the problem of a different scale being applied to the horizontal and vertical axes, respectively, as typically done in a projected coordinate system like UTM.

To enable direct comparison of the NeRF and MVS point clouds with the reference point cloud, the outputs were subsequently transformed back into the global CRS using the inverse transformation parameters. This approach provides an elegant way of avoiding transformation errors, ensures consistency across all datasets, and eliminates the need for an additional registration step (e.g., via an Iterative Closest Point (ICP) algorithm) (Besl and McKay, 1992). The success of the bundle adjustment was confirmed by the low positional error of the markers (avg. 7 mm / 0.2 px) and the stability of the internal orientation parameters (average RMSE reprojection error over all tie points = 0.2 px).

4.3 MVS point cloud generation

The MVS point cloud was generated using Agisoft Metashape. In contrast to the reference point cloud (Section 4.1), the images were not processed at full resolution. Instead, one level of image-pyramid downsampling was applied (Quality = high; 5 MP) to enable a meaningful comparison to the NeRF-derived point cloud, which also employed a one-level downsampling. For the depth-map generation, the mild filtering option was used. To ensure that the observed differences from the NeRF point cloud were not caused by this filtering, the MVS reconstruction was repeated without filtering. The resulting difference in point count was only 0.2%, which was considered negligible. Consequently, all subsequent analyzes were performed using the filtered point cloud, as this represents the default variant used in the software. After processing, the point cloud was exported, transformed into the global CRS, and cropped to the study area.

4.4 NeRF point cloud generation

For training and generation, the PyTorch-based framework Nerfstudio (V1.1.5) (Tancik et al., 2023) was employed, which can directly import the topocentric camera poses and internal parameters exported from Metashape after bundle adjustment. All images were downsampled once to enable training of the scene on an NVIDIA RTX4090 with 24GB VRAM. For training, the Nerfacto model with default parameters was selected, as it integrates advances from MipNeRF360 (Barron et al., 2022), NeRF— (Wang et al., 2021), Instant-NGP (Müller et al., 2022), NeRF-W (Martin-Brualla et al., 2021) and Ref-NeRF (Verbin et al., 2022). The camera optimizer implemented in Nerfacto was disabled because high-quality internal and external camera parameters were already available, and to maintain comparability with the MVS approach. The training was performed with 30k iterations, as this is the default value used by the model. Although the images were acquired exclusively from above the study area, the resulting peak signal-to-noise ratio (PSNR) of 20.02, structural similarity index (SSIM) of 0.46 and learned perceptual image patch similarity (LPIPS) of 0.63 indicate an acceptable NeRF reconstruction quality of the scene, when compared to the results of Nerfacto with the *Nerfstudio Dataset* presented in Tancik et al. (2023).

After training, a point cloud with 50 million points was created and subsequently filtered using the statistical outlier removal

(SOR) filter, implemented in Open3D (Zhou et al., 2018), with a standard deviation setting (std) of 2.0σ , and exported in the local topocentric CRS, where the study area was manually defined using an oriented bounding box (OBB). In addition, the *Noise filter*, integrated in CloudCompare (2025), was applied that locally fits a plane around each point and removes points whose deviation from the fitted plane exceeds a standard deviation of 1σ relative to the neighboring reprojection error. After filtering and cropping, the point cloud consisted of 18 million points. Thus, the NeRF point cloud contains more than four times the number of points in the MVS point cloud, despite using the same input data. This oversampling is an inherent advantage of NeRFs, although the interpretation of individual points should therefore be treated with caution.

4.5 3DGS point cloud generation

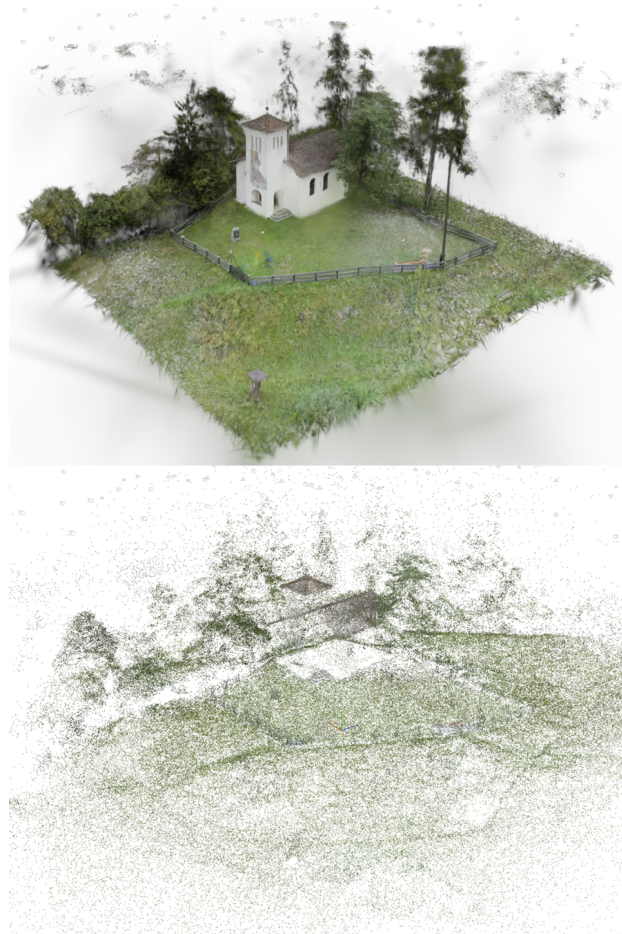


Figure 4. Comparison between the 2D view synthesis and the resulting point cloud of the scene in 3DGS.

The 3DGS model was trained using the software Postshot (V1.0.110). To ensure compatibility, interior and exterior orientation parameters were converted into the RealityScan file-format and imported with the sparse SfM point cloud and original imagery. Training was conducted using the 'Splat3' architecture for 30k iterations. To maintain consistency across the experimental setup, images were downsampled to match the resolution of the other two methods. Additionally, a Region of Interest (ROI) was defined to constrain the optimization process strictly to the study area.

Although the 3DGS approach yielded the most visually compelling results in terms of 2D novel view synthesis (SSIM:

0.567 for 3DGS vs. 0.46 for NeRF), as illustrated in Figure 4, it proved unsuitable for dense geometric reconstruction. This limitation stems from the inherent nature of 3DGS, where the resulting set of Gaussian primitives, approximately 1 million in this study, remains closer in magnitude to the initial sparse SfM points (160,000) than to a dense MVS point cloud. These observations align with findings by Clini et al. (2024) and Winawarter et al. (2025), who both noted low point densities in similar scenarios. A formal geometric evaluation of the 3DGS-derived points was omitted, as the method does not produce a continuous dense surface, rendering a comparative spatial analysis statistically non-significant for the scope of this research.

5. Evaluation

In this section, the dense point clouds are compared with respect to their processing times, geometric accuracy and reconstruction completeness. As a reference, the combined TLS/MVS point cloud presented in Section 4.1 was used. All datasets were cropped to the same ROI shown in Figure 2, which covers an area of approximately $20 \times 15 \text{ m}^2$. To evaluate the reconstruction quality of the chapel separately, this building was additionally extracted, yielding two datasets on which the following analyses were performed. In order to use the algorithms more efficiently, the reference point cloud was subsampled using a voxel size of 5 mm. This resolution is sufficient for the M3C2 diameter scale described below to accurately represent the scene as a reference point cloud and to calculate accurate normal vectors for the distance calculation. The NeRF point cloud was also subsampled to match the number of points of the MVS point cloud (in the following called NeRF-sub), to investigate whether the results were driven solely by a higher point count.

First, the processing time can be evaluated based on the hardware described in Section 4.4. The generation of the point cloud using MVS required 13 minutes, while training the NeRF model and exporting the corresponding point cloud took approximately 20 minutes, indicating that no substantial difference in processing time is observed between the two approaches.

Geometric accuracy was assessed by applying the M3C2 algorithm (Lague et al., 2013). For the M3C2 distance computation, the normal estimation scale (D) was set to 0.1 m, defining the orientation of the distance measurement. The projection scale (d), which determines the cylindrical neighborhood used for averaging points along the normal direction, was also set to 0.1 m.

The M3C2 distances also form the basis for the completeness analysis of the point clouds. Following the procedure described by Seitz et al. (2006), completeness is quantified as the proportion of reference points G that lie within a tolerance distance d of the reconstructed point cloud R . We adjusted the value of d , which was previously suggested to be 0.2 m by Hermann et al. (2024) for UAV datasets, to $\pm 0.1 \text{ m}$, as our GSD was very low for a UAV-based photogrammetric flight. Because the study area also contains large amounts of vegetation, the threshold was extended to $\pm 0.3 \text{ m}$ in a second analysis. The rationale for this choice was that the exact position of a tree branch is less important than the presence of a reconstructed point. For each of the five reconstructed point clouds R_i , the M3C2 distances were computed between the core points G_p and R_i . Points whose distance exceeded the tolerance d

were discarded and completeness was defined as the ratio of the remaining inlier reference points $G_{p < d}$ to the total number of reference points G_p .

Dataset	Method	Points [$\cdot 10^6$]
Chapel	Reference	85.1
	Reference-sub	13.8
	MVS	1.6
	NeRF	8.9
	NeRF-sub	1.6
Scene	Reference	139.1
	Reference-sub	31.8
	MVS	4.1
	NeRF	17.9
	3DGS	1.0

Table 1. Number of points obtained for each method. The "-sub" suffix marks all point clouds that have been subsampled to an voxel size of 5 mm.

The subsampled TLS reference point cloud G_p consisted of roughly 32 million points, and 14 million points for the detailed chapel subset, representing a substantially higher point count than any of the reconstructed clouds R_i (see Table 1). For each of the four reconstructed point clouds R_i , M3C2-distances were computed and completeness were calculated based on these.

6. Results and Discussion

This section presents and discusses the results of the accuracy and completeness analyses. Table 2 provides an overview of the numerical values. The presentation of results first focuses on the chapel subset, which represents a building with well-defined structures and surfaces, and subsequently on the entire study area to assess performance especially over vegetated regions.

With respect to the M3C2 distances relative to the reference point cloud, the chapel was successfully reconstructed in both datasets, as the mean deviation was 6 mm for MVS and 0 mm for NeRF, which both lie well below the flight ground sampling distance (GSD) of the non scaled images, so that the difference can be attributed to random noise. The mean value can be interpreted as a measure of georeferencing accuracy, which confirms the correct calculation and application of the workflow used. Based on the standard deviation of the distances, MVS achieved a higher geometric accuracy for the chapel (0.09 m) compared to the NeRF reconstruction (0.12 m), while the NeRF point cloud exhibited a 6% higher completeness. Although this difference appears moderate in absolute terms, Figures 5, 6 and 7 clearly show that NeRF covers substantially more areas. This can be attributed to regions that were not reconstructed at all in MVS but appear only sparsely represented in the NeRF point cloud. This observation is further supported by the fact that, after subsampling the reference cloud to a GSD of 0.1 m and repeating the completeness analysis, the completeness value of MVS remained nearly constant (67.6%) whereas NeRF increased to 76.8%, a difference exceeding 9 %. These sparsely reconstructed areas in the NeRF point cloud also exhibit the lowest geometric accuracy (see Figure 5).

To test whether the higher completeness of the NeRF point cloud was solely due to its higher point count, the NeRF point cloud was randomly subsampled to match the MVS point count (NeRF-sub with 1.6 million points). Even then, NeRF still achieved 2.6% higher completeness, showing the expected

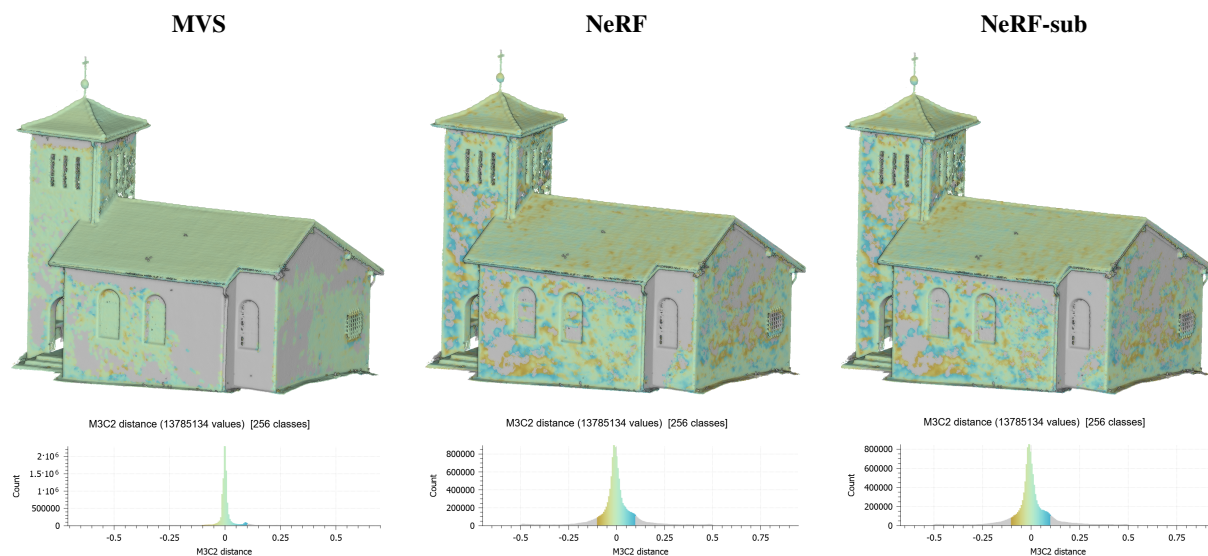


Figure 5. M3C2 distances (chapel scene) of the image-based point clouds to the reference point cloud. Grey areas have not been reconstructed, whereas the other colors represent distances and can be compared with the histogram representations at the bottom.

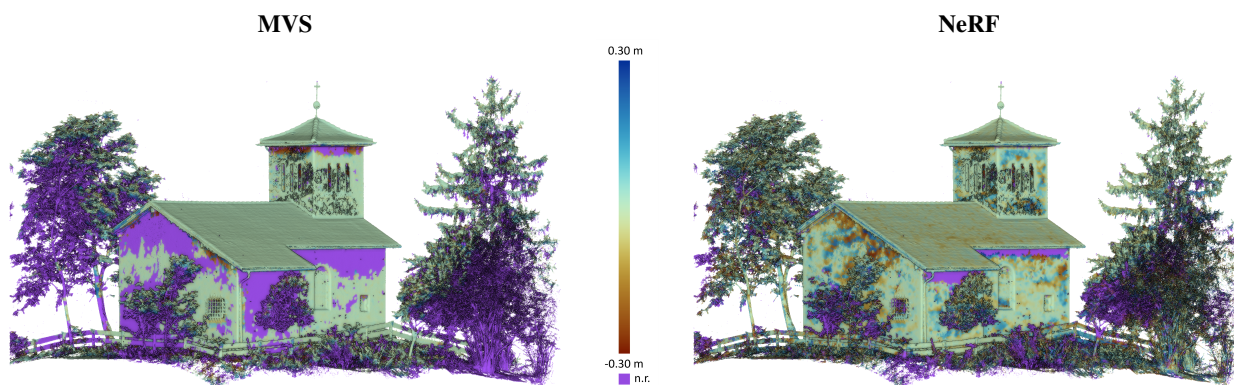


Figure 6. M3C2 distances (whole scene) of the image-based point clouds to the reference point cloud. Areas that could not be reconstructed (n.r.) are shown in purple for better clarity.



Figure 7. Detailed comparison of the completeness of the different methods. Green areas could be reconstructed by both methods, whereas red areas were only reconstructed by the NeRF point cloud. Grey areas could not be reconstructed successfully by either method.

methodological advantages. To demonstrate the effect of the filters on the point cloud, the NeRF point cloud was also analyzed without applying the SOR and noise filters (NeRF-raw). As expected, the M3C2 accuracy (standard deviation) value increases to 0.23 m while the completeness remains roughly the same (Table 2). However, the point cloud visually appears very noisy, which may be undesirable.

A similar trend is observed for the entire study area. MVS again achieved higher accuracy (0.11 m) compared to NeRF (0.15 m, Table 2). This difference is primarily due to the grass areas, which are cleanly represented in the MVS point cloud but appear to be heavily noisy in the NeRF reconstruction, a behavior also reported in previous studies, e.g. in Winiwarter et al. (2025). However, the NeRF point cloud represents the full study area 20% more complete. In particular, the lower vegetation layers of the trees are much better reconstructed, which is consistent with the findings of Tian et al. (2025).

In summary, the NeRF reconstruction in both datasets (the full scene and the chapel subset) has slightly lower accuracy but significantly higher completeness. Especially in areas where MVS cannot capture any points at all, the NeRF reconstruction is able to provide information about the surface location, which enables a better reconstruction of e.g. the chapel walls or the

vegetation.

Dataset	Method	Mean [m]	Std [m]	Compl. [%]
Chapel d ± 0.1	MVS	0.006	0.090	68.2
	NeRF	0.000	0.123	74.4
	NeRF-sub	0.003	0.124	70.8
Scene d ± 0.3	MVS	0.016	0.108	66.6
	NeRF	-0.012	0.153	86.7
	NeRF-raw	-0.013	0.227	84.7

Table 2. Mean and standard deviation (Std) [1σ] of the M3C2 distances and the completeness evaluation for the chapel and the whole study area. For estimating the completeness, a distance threshold of ± 0.1 m and ± 0.3 m were used.

7. Conclusion and Outlook

This study investigated the reconstruction quality using two image-based methods in a challenging environment. Dense point clouds were generated using both MVS and NeRF methods and subsequently compared relative to a reference point cloud with respect to geometric accuracy and completeness. To ensure comparability, identical internal and external camera orientation parameters and the same set of 105 nadir and oblique images, downsampled to the same resolution (5 MP), were used. To obtain a reference of superior accuracy, the study area was additionally surveyed from 16 TLS stations, except for the roof areas, which were not accessible from the ground. These roof areas were supplemented with an image-based point cloud using oblique imagery at maximum resolution. A workflow was successfully developed that enables the use of NeRFs in a realistic photogrammetric, georeferenced scenario, particularly in areas where MVS cannot reconstruct points. In addition, it has been shown that NeRF point clouds can be used with classical point cloud-based evaluation methods, such as accuracy assessments via M3C2 and completeness assessments.

The results show that for both the building and the surrounding vegetation, NeRF-based point clouds exhibit lower geometric accuracy than MVS but achieved a higher degree of completeness. This may be explained by the different point generation mechanisms, where MVS produces discrete pixel-based points, whereas NeRF derives points from a continuous density function. The NeRF method can be advantageous in complex scenes and shows parallels to full-waveform analysis in laser scanning (Mallet and Bretar, 2009), where the received laser signal is comparable to the signal of the volumetric density along a ray in NeRFs.

Compared to the TLS point cloud, both image-based methods perform respectably. Future work should focus on strategies to further improve the quality of NeRF point clouds and on methods for integrating the results of both approaches to obtain reconstructions that are accurate and more complete than those produced by classical photogrammetric algorithms. This combined procedure can then also be tested in other areas, including high-alpine geomorphic study sites.

References

Agisoft LLC, 2025. Terrestrial laser scans processing in Metashape 2.2 - Agisoft Helpdesk Portal. <https://agisoft.freshdesk.com/support/solutions/articles/31000168474-terrestrial-laser-scans-processing-in-metashape-2-2> (25.09.2025).

Aicardi, I., Chiabrando, F., Maria Lingua, A., Noardo, F., 2018. Recent trends in cultural heritage 3D survey: The photogrammetric computer vision approach. *Journal of Cultural Heritage*, 32, 257–266. <https://doi.org/10.1016/j.culher.2017.11.006>.

Barron, J. T., Mildenhall, B., Verbin, D., Srinivasan, P. P., Hedman, P., 2022. Mip-NeRF 360: Unbounded Anti-Aliased Neural Radiance Fields. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5470–5479. <https://doi.org/10.48550/ARXIV.2111.12077>.

Besl, P., McKay, N. D., 1992. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2), 239–256. <https://doi.org/10.1109/34.121791>.

Billi, D., Caroti, G., Piemonte, A., 2025. Metric Error Assessment Regarding Geometric 3D Reconstruction of Transparent Surfaces via SfM Enhanced by 2D and 3D Gaussian Splatting. *Sensors*, 25(14), 4410. <https://doi.org/10.3390/s25144410>.

CANON INC., 2006. EF LENS WORK III - the Eyes of EOS. https://ia601801.us.archive.org/28/items/canon-ef_lens_work_iii-the_eye_of_eos-8e-2006 (25.09.2025).

Clini, P., Nespeca, R., Angeloni, R., Coppetta, L., 2024. 3D representation of Architectural Heritage: a comparative analysis of NeRF, Gaussian Splatting, and SfM-MVS reconstructions using low-cost sensors. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-2/W8-2024, 93–99. <https://doi.org/10.5194/isprs-archives-XLVIII-2-W8-2024-93-2024>.

CloudCompare, 2025. CloudCompare (version 2.13.2). <https://www.cloudcompare.org> (25.09.2025).

Condorelli, F., Rinaudo, F., Salvatore, F., Tagliaventi, S., 2021. A comparison between 3d reconstruction using nerf neural networks and mvs algorithms on cultural heritage images. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B2-2021, 565–570. <https://doi.org/10.5194/isprs-archives-XLIII-B2-2021-565-2021>.

DJI Enterprise, 2021. Aerial Surveying Just Got Smart. <https://enterprise-insights.dji.com/blog/smart-oblique-capture> (25.09.2025).

DJI Enterprise, 2025. Techn. Data DJI Mavic 3M. <https://enterprise.dji.com/en/mavic-3-m/specs> (25.09.2025).

Fischler, M. A., Bolles, R. C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24, 395. <https://doi.org/10.1145/358669.358692>.

Furukawa, Y., Hernández, C., 2015. Multi-View Stereo: A Tutorial. *Foundations and Trends® in Computer Graphics and Vision*, 9(1), 1–148. <https://doi.org/10.1561/06000000052>.

Furukawa, Y., Ponce, J., 2007. Accurate, Dense, and Robust Multi-View Stereopsis. *2007 IEEE Conference on Computer Vision and Pattern Recognition*, 1–8. <https://doi.org/10.1109/CVPR.2007.383246>.

Galliani, S., Lasinger, K., Schindler, K., 2015. Massively Parallel Multiview Stereopsis by Surface Normal Diffusion. *2015 IEEE International Conference on Computer Vision (ICCV)*, 873–881. <https://doi.org/10.1109/iccv.2015.106>.

- Gonizzi Barsanti, S., Guagliano, M., Rossi, A., 2022. 3D Reality-Based Survey and Retopology for Structural Analysis of Cultural Heritage. *Sensors*, 22(24), 9593. <https://doi.org/10.3390/s22249593>.
- Hermann, M., Weinmann, M., Nex, F., Stathopoulou, E., Remondino, F., Jutzi, B., Ruf, B., 2024. Depth estimation and 3D reconstruction from UAV-borne imagery: Evaluation on the UseGeo dataset. *ISPRS Open Journal of Photogrammetry and Remote Sensing*, 13. <https://doi.org/10.1016/j.ophoto.2024.100065>.
- Huggel, C., Clague, J. J., Korup, O., 2012. Is climate change responsible for changing landslide activity in high mountains? *Earth Surface Processes and Landforms*, 37(1), 77–91. <https://doi.org/10.1002/esp.2223>.
- Irwin, L. A. K., Coops, N. C., Anders, K., Mandlbürger, G., Winiwarter, L., 2025. Light detection and ranging of natural systems. *Nature Reviews Methods Primers*, 5(1), 76. <https://doi.org/10.1038/s43586-025-00446-3>.
- Kerbl, B., Kopanas, G., Leimkuehler, T., Drettakis, G., 2023. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Transactions on Graphics*, 42(4), 1–14. <https://doi.org/10.1145/3592433>.
- Knapitsch, A., Park, J., Zhou, Q.-Y., Koltun, V., 2017. Tanks and Temples: Benchmarking Large-Scale Scene Reconstruction. *ACM Transactions on Graphics*, 36(4). <https://www.tanksandtemples.org/>.
- Lague, D., Brodu, N., Leroux, J., 2013. Accurate 3D comparison of complex topography with terrestrial laser scanner: Application to the Rangitikei canyon (N-Z). *ISPRS Journal of Photogrammetry and Remote Sensing*, 82, 10–26. <https://doi.org/10.1016/j.isprsjprs.2013.04.009>.
- Lowe, D. G., 2004. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2), 91–110. <https://doi.org/10.1023/B:VISI.0000029664.99615.94>.
- Luhmann, T., 2010. Erweiterte Verfahren zur geometrischen Kamerakalibrierung in der Nahbereichsphotogrammetrie. *Habilitation TU Dresden. (German Geodetic Commission), C 645*.
- Mallet, C., Bretar, F., 2009. Full-waveform topographic lidar: State-of-the-art. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(1), 1–16. <https://doi.org/10.1016/j.isprsjprs.2008.09.007>.
- Martin-Brualla, R., Radwan, N., Sajjadi, M. S. M., Barron, J. T., Dosovitskiy, A., Duckworth, D., 2021. NeRF in the Wild: Neural Radiance Fields for Unconstrained Photo Collections. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7210–7219. <https://doi.org/10.48550/ARXIV.2008.02268>.
- Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., Ng, R., 2021. NeRF: representing scenes as neural radiance fields for view synthesis. *Commun. ACM*, 65(1). <https://doi.org/10.1145/3503250>.
- Mohren, J., Schulze, M., 2024. Automated cleaning of tie point clouds following USGS guidelines in Agisoft Metashape professional (ver. 2.1.0). *MethodsX*, 12. <https://doi.org/10.1016/j.mex.2024.102679>.
- Müller, T., Evans, A., Schied, C., Keller, A., 2022. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics*, 41(4), 1–15. <https://doi.org/10.1145/3528223.3530127>.
- Niemeyer, M., Mescheder, L., Oechsle, M., Geiger, A., 2020. Differentiable Volumetric Rendering: Learning Implicit 3D Representations Without 3D Supervision. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2020)*, 3501–3512. <https://doi.org/10.1109/CVPR42600.2020.00356>.
- Oechsle, M., Peng, S., Geiger, A., 2021. UNISURF: Unifying Neural Implicit Surfaces and Radiance Fields for Multi-View Reconstruction. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 5589–5599. <https://doi.org/10.48550/ARXIV.2104.10078>.
- Over, J.-S. R., Ritchie, A. C., Kranenburg, C. J., Brown, J. A., Buscombe, D., Noble, T., Sherwood, C. R., Warrick, J. A., Wernette, P. A., 2021. Processing coastal imagery with Agisoft Metashape Professional Edition, version 1.6—Structure from motion workflow documentation: U.S. Geological Survey Open-File Report, 1039. <https://doi.org/10.3133/ofr20211039>.
- Petrovska, I., Jutzi, B., 2025. Seeing beyond vegetation: A comparative occlusion analysis between Multi-View Stereo, Neural Radiance Fields and Gaussian Splatting for 3D reconstruction. *ISPRS Open Journal of Photogrammetry and Remote Sensing*, 16, 100089. <https://doi.org/10.1016/j.ophoto.2025.100089>.
- Remondino, F., Karami, A., Yan, Z., Mazzacca, G., Rigon, S., Qin, R., 2023. A Critical Analysis of NeRF-Based 3D Reconstruction. *Remote Sensing*, 15(14). <https://doi.org/10.3390/rs15143585>.
- Riegl, 2025a. RiSCAN PRO: Terrestrial Laser Scanning Software. <https://www.riegl.com/de-oesterreich/products/detail/riscan-pro> (25.09.2025).
- Riegl, 2025b. Terrestrial Laser Scanning - Riegl VZ-600i. <https://www.riegl.com/de-oesterreich/products/detail/riegl-vz-600i> (25.09.2025).
- Rietdorf, A., 2005. Automatisierte Auswertung und Kalibrierung von scannenden Messsystemen mit tachymetrischem Messprinzip. *PhD thesis TU Berlin. (German Geodetic Commission), C 582*.
- Schönberger, J. L., Frahm, J.-M., 2016. Structure-from-Motion Revisited. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4104–4113. <https://doi.org/10.1109/CVPR.2016.445>.
- Schönberger, J. L., Zheng, E., Pollefeys, M., Frahm, J.-M., 2016. Pixelwise View Selection for Unstructured Multi-View Stereo. *European Conference on Computer Vision (ECCV)*. https://doi.org/10.1007/978-3-319-46487-9_31.
- Schulte, F., Andreas Mair, Perpetual Hope Akwensi, Leon Bührle, Lukas Winiwarter, 2026. Bilder statt LiDAR? Potenzial neuronaler Netze für die hochauflösende 3D-Erfassung komplexer Waldstrukturen. 46. *Wissenschaftlich-Technische Jahrestagung der DGPF in Darmstadt*.

- Schulte, F., Brezovsky, M., Günthner, A., Jutzi, B., Mandlbürger, G., Winiwarter, L., 2025. Simulation and validation of underwater scenes for two-media optical 3D reconstruction. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-2/W10-2025, 271–278. <https://doi.org/10.5194/isprs-archives-xlvi-2-w10-2025-271-2025>.
- Seitz, S., Curless, B., Diebel, J., Scharstein, D., Szeliski, R., 2006. A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms. *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 1 (CVPR'06)*, 519–528. <https://doi.org/10.1109/CVPR.2006.19>.
- Tancik, M., Weber, E., Ng, E., Li, R., Yi, B., Wang, T., Kristoffersen, A., Austin, J., Salahi, K., Ahuja, A., McAllister, D., Kerr, J., Kanazawa, A., 2023. Nerfstudio: A Modular Framework for Neural Radiance Field Development. *SIGGRAPH '23: Special Interest Group on Computer Graphics and Interactive Techniques Conference*, 1–12. <https://doi.org/10.1145/3588432.3591516>.
- Teruggi, S., Grilli, E., Fassi, F., Remondino, F., 2021. 3D surveying, semantic enrichment and virtual access of large cultural heritage. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, VIII-M-1-2021, 155–162. <https://doi.org/10.5194/isprs-annals-VIII-M-1-2021-155-2021>.
- Tian, G., Chen, C., Huang, H., 2025. Comparative Analysis of Novel View Synthesis and Photogrammetry for 3D Forest Stand Reconstruction and Extraction of Individual Tree Parameters. *Remote Sensing*, 17(9). <https://doi.org/10.3390/rs17091520>.
- Triggs, B., McLauchlan, P. F., Hartley, R. I., Fitzgibbon, A. W., 2000. Bundle adjustment - a modern synthesis. B. Triggs, A. Zisserman, R. Szeliski (eds), *Vision Algorithms: Theory and Practice. IWVA 1999. Lecture Notes in Computer Science*, 1883, Springer Berlin Heidelberg, 298–372. https://doi.org/10.1007/3-540-44480-7_21.
- Ullrich, A., Fürst, C., 2017. Vollautomatischer Ansatz für die Onboard- Datenregistrierung im terrestrischen Laserscanning. *Terrestrisches Laserscanning 2017 (TLS 2017) - Beiträge zum 165. DVW-Seminar am 11. und 12. Dezember 2017 in Fulda - Schriftenreihe des DVW*, 88.
- Verbin, D., Hedman, P., Mildenhall, B., Zickler, T., Barron, J. T., Srinivasan, P. P., 2022. Ref-NeRF: Structured View-Dependent Appearance for Neural Radiance Fields. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5481–5490. <https://doi.org/10.48550/ARXIV.2112.03907>.
- Wang, Z., Wu, S., Xie, W., Chen, M., Prisacariu, V. A., 2021. NeRF- -: Neural Radiance Fields Without Known Camera Parameters. <https://doi.org/10.48550/ARXIV.2102.07064>.
- Wilkinson, M., Jones, R., Woods, C., Gilment, S., McCaffrey, K., Kokkalas, S., Long, J., 2016. A comparison of terrestrial laser scanning and structure-from-motion photogrammetry as methods for digital outcrop acquisition. *Geosphere*, 12(6), 1865–1880. <https://doi.org/10.1130/GES01342.1>.
- Winiwarter, L., Anders, K., Höfle, B., 2021. M3C2-EP: Pushing the limits of 3D topographic point cloud change detection by error propagation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 178, 240–258. <https://doi.org/10.1016/j.isprsjprs.2021.06.011>.
- Winiwarter, L., Schulte, F., Wang, J., Zhang, Q., Anders, K., Jutzi, B., 2025. Assessing the Potential of Neural Radiance Fields and Gaussian Splatting for Change Detection and Change Quantification. *Proceedings of the 6th Joint International Symposium on Deformation Monitoring - JISDM 2025*. <https://doi.org/10.5445/IR/1000180533>.
- Xiang, L., Zheng, H., Deng, B., Yin, H., 2025. Geometric Prior-Guided Neural Implicit Surface Reconstruction in the Wild. *arXiv*. <https://doi.org/10.48550/ARXIV.2505.07373>.
- Yan, Z., Mazzacca, G., Rigon, S., Farella, E. M., Trybala, P., Remondino, F., 2023. NERFBK: A holistic dataset for benchmarking nerf-based 3d reconstruction. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-1/W3-2023, 219–226. <https://doi.org/10.5194/isprs-archives-XLVIII-1-W3-2023-219-2023>.
- Yan, Z., Padkan, N., Trybala, P., Farella, E. M., Remondino, F., 2025. Learning-Based 3D Reconstruction Methods for Non-Collaborative Surfaces—A Metrological Evaluation. *Metrology*, 5(2), 20. <https://doi.org/10.3390/metrology5020020>.
- Yariv, L., Gu, J., Kasten, Y., Lipman, Y., 2021. Volume Rendering of Neural Implicit Surfaces. *Advances in Neural Information Processing Systems*, 34, 4805–4815. <https://doi.org/10.48550/ARXIV.2106.12052>.
- Zhou, Q.-Y., Park, J., Koltun, V., 2018. Open3D: A Modern Library for 3D Data Processing. *arXiv:1801.09847*. <https://doi.org/10.48550/arXiv.1801.09847>.