Simulation and validation of underwater scenes for two-media optical 3D reconstruction

Frederik Schulte[1], Markus Brezovsky[2], Anatol Günthner[3], Boris Jutzi[3], Gottfried Mandlburger[2], Lukas Winiwarter[1]

[1] Unit of Geometry and Surveying, Faculty of Engineering Sciences, University of Innsbruck, Innsbruck, Austria
(frederik.schulte, lukas.winiwarter)@uibk.ac.at
[2] Department of Geodesy and Geoinformation, TU Wien, Vienna, Austria -
(markus.brezovsky, gottfried.mandlburger)@tuwien.ac.at
[3] Institute of Photogrammetry and Remote Sensing, Karlsruhe Institute of Technology, Karlsruhe, Germany -
(anatol.guenthner, boris.jutzi)@kit.edu

Keywords: Simulation, multimedia photogrammetry, ray tracing, validation.

Abstract

Optical 3D reconstruction in environments with complex light paths such as water-air surface interactions continues to be challenging, particularly due to the inherent refraction effects. These effects compromise assumptions taken in standard photogrammetric methods like the traditional Multi-View Stereo and newer approaches like Neural Radiance Fields (NeRFs). Addressing these limitations is critical for monitoring coastal and riparian ecosystems, for flood-risk modeling, as climate change intensifies river flooding, and in general to satisfy increasing demands for 3D topo-bathymetric data. To evaluate models explicitly built to consider a change in refractivity along the image ray, simulation can be employed. In this study, we present a simulation and validation framework designed to investigate these challenges by synthesizing controlled water scenes with artificial camera trajectories and evaluating them with 2D and 3D (Cloud-to-Mesh, completeness) metrics. For that, a total of 130 images with a resolution of 1024 × 768 pixels were simulated to model both a water-free scene and a submerged scene. The results indicate that refractive effects must be explicitly accounted for, as a water depth of 3.5 m led to errors on the order of 1 m, when refraction was not taken into account. Furthermore, NeRFs proved to be particularly well suited for 3D analysis of photo-bathymetric surveys, achieving a completeness that was 21 % higher than conventional MVS methods. The simulation workflow is particularly beneficial for the development and testing of specialized NeRF-variants designed to better account for the complexities introduced by refraction at air-water interface.

1. Introduction

Anthropogenic climate change and concomitant global temperature rise are profoundly affecting ecosystems and represent one of the foremost societal challenges of our era. These effects are particularly pronounced in aquatic environments, where elevated temperatures promote the occurrence of intense precipitation events and induce shifts in submerged biotic communities. Underwater photogrammetry has demonstrated its utility for high-resolution mapping of shallow-water bathymetry. Concurrently, the surveillance of submerged vegetation has assumed growing importance, as the distribution and condition of macrophyte assemblages serve as sensitive indicators of climatic perturbations (Lind et al., 2022). To date, efficient large-scale surveys of submerged vegetation have relied predominantly on active laser-based remote sensing techniques utilizing a green laser. However, recent advances in machine learning, specifically Neural Radiance Fields (NeRFs), hold promise for developing purely image-based, cost-effective methodologies for three-dimensional mapping of underwater habitats, where Multi-View Stereo (MVS) struggles. To improve new, refraction-aware NeRF models, simulations can be used to create synthetic image datasets, thereby enabling rigorous validation of refractive parameter estimation and reconstruction accuracy under controlled scenarios. Upon completion of the simulation-based testing phase, these models can be employed to analyze real-world scenarios.

1.1 BathyNeRF

BathyNeRF, a transnational research project between the authors' institutions, focuses primarily on enhancing existing methodologies for the detection and three-dimensional modeling of underwater topography and submerged vegetation with radiance fields. Its primary objectives are, firstly, to refine NeRF-based algorithms for reconstructing underwater topography and underwater plants from above-water image acquisitions, and secondly, to perform a rigorous quantitative evaluation against established benchmarks, namely MVS photo-bathymetry and airborne laser bathymetry. A central research challenge involves the incorporation of refraction-induced phenomena into the NeRF framework to ensure faithful reconstruction of subaqueous environments. During the project's initial phase, empirical data acquisition was conducted at the Pielach River study site (Mandlburger et al., 2025). Additionally, NeRFrac code base (Zhan et al., 2023) was modified to integrate refractive-index corrections (Guenthner et al., 2025) and to enable post-training point cloud export (Brezovsky et al., 2025). The simulation framework described in this article constitutes a critical component of this first phase, facilitating controlled experiments to validate and calibrate the extended NeRF algorithms prior to their application in real-world aquatic scenarios.

2. Methodology

In this section, the fundamentals of the employed reconstruction techniques are first presented. Section 2.1 provides an overview of the Structure-from-Motion (SfM) workflow. Based on these results, a dense point cloud is generated using MVS in Section 2.2 and a radiance field-based approach with NeRFs in Section 2.3.

2.1 SfM

SfM is a photogrammetric technique that reconstructs 3D scenes from 2D images by simultaneously estimating camera poses and the tie point cloud (Schönberger and Frahm, 2016). The process begins with the detection of distinctive features, which are commonly extracted by the Scale Invariant Feature Transform (SIFT) algorithm (Lowe, 2004) or variants thereof. These features are then matched across stereo image pairs using the outlier-robust Random Sample Consensus (RANSAC) algorithm (Fischler and Bolles, 1981). From the resulting correspondences, the relative orientations of the camera positions are estimated. Finally, a bundle adjustment is performed to refine all camera parameters and 3D point locations by minimizing the overall reprojection error, thereby enhancing both the camera calibration and the accuracy of the tie points (Triggs et al., 2000).

2.2 MVS

MVS is a framework that builds on the SfM-workflow and computes a dense point cloud from the sparse tie point cloud together with the cameras' interior and exterior orientation parameters (Schönberger et al., 2016). MVS first identifies corresponding image points across overlapping views of the same scene and uses these correspondences to estimate depth values for each pixel in every picture. These depth maps are then matched and fused to produce a complete, dense 3D point cloud of the scene (Furukawa and Hernández, 2015). On smooth, well-textured surfaces, these algorithms perform robustly and can rival the results of much more expensive laser scanning systems. However, under poor lighting conditions, particularly on textureless, reflective, or refractive surfaces, the quality of the reconstruction deteriorates significantly and may become unusable.

2.3 NeRF

NeRFs are a deep-learning technique for implicitly representing 3D scenes from calibrated photographs. Given a set of input images with known exterior and interior camera parameters, a neural network is trained to synthesize novel views from arbitrary directions. Unlike traditional geodetic methods that explicitly store 3D points, NeRFs encode the entire scene within the network's learned weights.

In the Vanilla-NeRF formulation (Mildenhall et al., 2021), the input to the network is a 5D vector comprising a 3D spatial location (x, y, z) and a 2D viewing direction (θ, ϕ). Along each camera ray intersecting the scene, these vectors are sampled. For each sample the model yields an RGB-color (R, G, B) alongside a volume density (σ). Synthetic images are rendered during training by numerically integrating the colors and densities along rays according to classical volume-rendering principles (Yariv et al., 2021), where σ represents the differential probability of the ray terminating at that point. The network is optimized by minimizing the rendering loss, which is the difference between these synthetic images and the actual photographs, leveraging the differentiability of the volume-rendering equation to adjust the multilayer perceptron's weights until the scene is faithfully reconstructed.

To produce a 3D point cloud comparable to that obtained via classical MVS, a post-processing step extracts surface geometry from the learned radiance field. During point cloud extraction, camera rays from the withheld test set are traced until the predicted volume density exceeds a fixed threshold, and the corresponding RGB value at that point is recorded as a surface sample (Oechsle et al., 2021).

3. Simulation

This section outlines the simulation framework shown in Figure 1. Section 3.1 covers the generation of the reference mesh, followed by the flight-planning workflow in Section 3.2, and the Blender-based simulation implementation in Section 3.3 (Blender Development Team, 2024).

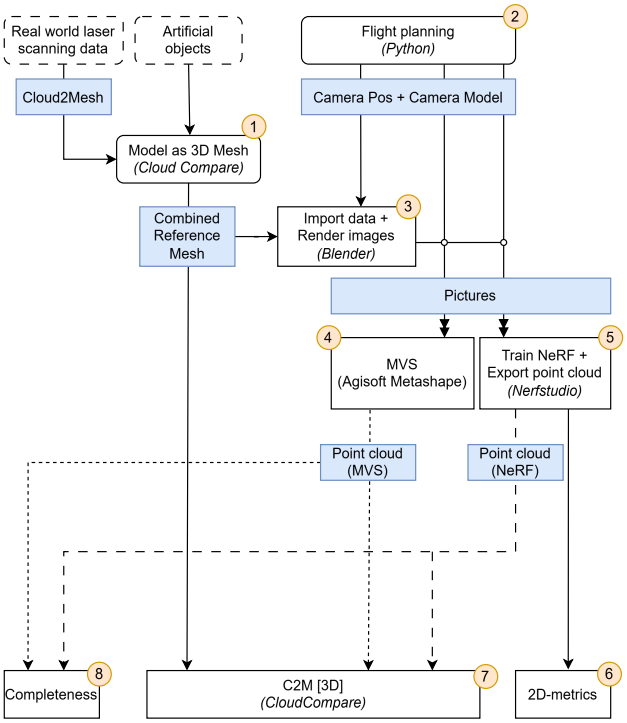

Figure 1. Simulation framework for 3D underwater reconstruction.

3.1 Create combined reference mesh

Prior to validating the MVS and NeRF point clouds, an initial reference dataset must be established. This dataset should be available as a mesh so that the subsequently generated point clouds can be assessed via a Cloud-to-Mesh (C2M) comparison. Meshes can also easily be imported in the simulation software Blender. Simulations offer the major advantage here that both datasets derived from topographic surveys and purely synthetic datasets both can be employed. Real datasets have the benefit that simulation can be optimized for a specific area or future survey site. However, if no data are available or existing data lack sufficient accuracy or resolution, synthetic data can still be generated.

To illustrate this benefit, we first utilized a laser scanning dataset of the former Jamtal glacier valley (WGS84: 46.90° N, 10.17° E), located above 2000 m a.s.l. The site features a river channel that feeds into a broad, flat gravel bar, providing a varied yet representative terrain for our simulation. In order to render the laser scanning point cloud usable for further analyses, the dataset was converted into a mesh in CloudCompare

using a 2.5D Delaunay triangulation. The raw survey data, acquired by a DJI L2 LiDAR scanner, had a ground sampling distance (GSD) of approximately 4 cm and was colored with the internal L2-camera. To prevent artifacts during later ray tracing, a Laplacian smoothing algorithm was applied to the resulting mesh using CloudCompare (Version 2.13.2).

As the synthetic dataset, a freely downloadable ship model was selected[1]. Placing the ship underwater not only complements the thematic context of the scene but also provides an interesting structural object for analysis, as it exhibits both large-scale curvatures and fine details. Although the comparison of the reconstruction methods is not the primary focus of this study, it will enable comparative reconstruction analyses between MVS and NeRF approaches.

Finally, the combined reference mesh, incorporating both the laser-scaned terrain and the synthetic ship, can be exported as an fbx-file, preserving vertex coloration for subsequent simulation steps. This composite mesh serves as the foundational geometry for all further investigations. Ideally, the edge length of the mesh should be below the GSD of the simulated flight mission. Unfortunately, this was not achieved here, but will be remedied in future measurements by using terrestrial laser scan measurements with a higher resolution

3.2 Generation of flight path

To achieve the most realistic simulation possible, the camera positions and flight trajectory must resemble those of an actual UAV mission. In order to test different scenarios, a Python script was developed which calculates the exterior orientation parameters of the camera (X, Y, Z, ϕ, ω, κ). For this, the camera model must be defined. This comprises the focal length, the pixel pitch (i.e. the pixel spacing on the sensor), and the sensor resolution. Other intrinsic calibration parameters, such as the sensor's principal point and lens distortion coefficients, were not modeled. Table 1 lists the specifications used for this experiment. A low resolution camera sensor was chosen so

Parameter	Value
Focal length	35 mm
Pixel pitch	20 μm
Sensor resolution	1024 × 768 (0.8 MP)

Table 1. Virtual camera model parameters used in this study.

that the subsequent NeRF training can be performed on original resolution images, thereby facilitating a direct comparison with MVS. Training on large numbers of high-resolution images from conventional drone camera systems (e.g. Sony Alpha 7 IV: 32.7 MP, DJI Zenmuse P1: 44.7 MP, Phase One iXM: 100 MP) presents significant challenges even on state-of-the-art hardware (NVIDIA RTX 4090 - 24 GB VRAM). The resource requirements are so high that multi-resolution image pyramids are usually generated prior to training in order to reduce the effective resolution, hindering direct comparisons to the MVS results.

Next, the flight parameters must be specified. In this case these are the lateral and longitudinal overlap, the GSD and the number of flight strips (or total area covered). Previous investigations with NeRFs have shown that oblique images are essential for robust geometric reconstruction, so these were also implemented. As illustrated in Figure 3, in the central region above the ship, nadir and oblique images towards the point of interest (POI, located at the center of the ship) are acquired first. Subsequently, three additional strips surrounding the central region are simulated, in which only oblique shots towards the direction of the POI are taken. Initially, the DJI Smart-Oblique mode (DJI Enterprise, 2021) was employed, which simulates a penta-camera (nadir + four obliques) above the area of interest. However, the dimensions of the site were found to be too small too small for that mode, as the oblique frames would have missed the target area. Therefore, an adapted approach was used in which only a single oblique shot towards the direction of the POI is acquired at each position. Table 2 summarizes the flight parameters used. Using these parameters, two text files were

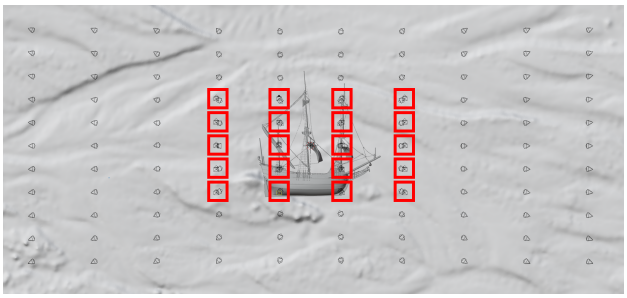

Figure 2. Camera positions on top of the untextured model in Blender. The positions that are used to take pictures in both nadir and oblique orientations are outlined in red.

Parameter	Value
Lateral overlap	60 %
Longitudinal overlap	80 %
GSD	3 cm
Number of central-region points	[4 x 5]
Number of outer oblique strips	3

Table 2. Flight parameters applied during the UAV mission.

generated: One containing the camera positions of the image sets and the other specifying the camera model. These files can be imported into both Blender (Version 4.3.2) and the photogrammetry software Agisoft Metashape (Version 2.2.1) used in this work. In the future, we plan to extend the script to enable the simulation of additional flight scenarios.

3.3 Simulation in Blender

In this section, the simulation with the open-source software Blender is described. Here, the 3D scene is converted into 2D image captures, thus simulating a photogrammetric mission. Further details on all settings can be found in the program documentation (Blender Development Team, 2024). First, the combined mesh from the real laser scan and the synthetic ship dataset are imported in the software. In addition, appropriate scene lighting must be chosen. To avoid additional reflection effects that would disturb the simulation, a luminous plane matching the scene dimensions was chosen, which illuminates the scene uniformly with an exposure simulating an overcast day. In general, the simulation was kept as simple as possible to specifically investigate the refractive influence of water. To investigate the effect of refraction on the evaluation models, a

[1] © user m23ldx0191. *Dutch ship* is licensed under CC BY 4.0. No changes were made. (26.05.2025) Available at: https://sketchfab.com/3d-models/dutch-ship-6e1a27f4db4c48e6adc53f6a3528ac9d

water body must be inserted into the scene. For this purpose, a rectangular water block with a depth of 3.5 m is added on top of the imported laser scan surface. This represents a realistic use case, since the image centers of all pictures lie on the plane with an elevation of 42.5 m above the terrain and thus the light rays spend only a small portion of their path underwater. Next, so-called BSDF (Bidirectional Scattering Distribution Function) shaders can be assigned to this water. These provide the mathematical basis for how light is scattered or refracted at a surface (Bartell et al., 1981). As inputs, an index of refraction of 1.333 and a color of (R = 0.8 / G = 0.9 / B = 1.0) are chosen to simulate realistic water properties. Furthermore, a damping coefficient (Volume Absorption) can be introduced, which, according to the exponential Beer–Lambert law, describes how strongly a light ray is absorbed in water. For this simulation, a value of 0.01 m^{-1} was used, corresponding to very clear water. Subsequently, the 130 camera positions (exterior orientation) with their associated camera model (interior orientation without calibration parameters) could be imported automatically via Blender's Python API. The camera's depth of field and motion blur were disabled to avoid introducing additional sources of error into the simulation. This simplification is not given in real photos, as there are always distorted images, which must be corrected afterwards by a camera calibration (Luhmann, 2010). Now, images must be rendered from the synthetic scene. For this, the Cycles engine, Blender's physically based path tracer was chosen, which follows rays from the camera and computes both direct and indirect illumination. To best simulate the scene, various settings must be configured, which are explained below: First, the number of samples, i.e. the number of paths per pixel in the final image, must be specified. An adaptive sampling scheme with a maximum of 4096 samples was selected. Once the noise threshold is reached, image quality is deemed sufficient and rendering terminates. The Max Bounces, i.e. the number of reflection, refraction, and scattering events per photon, were limited to twelve, which on one hand allows realistic light interactions and on the other hand keeps render times within practical bounds. The OptiX denoising algorithm (Chaitanya et al., 2017), a GPU-accelerated AI method that can reduce the rendering time of a high-resolution image without visible artifacts, was disabled to avoid stochastic deviations. For the same reason, the Compositing option was disabled too, so that Blender outputs the image produced directly by the Cycles engine without any color corrections or effects. In this way, refraction, which is the primary focus of this investigation, as well as reflection and absorption of the light rays can be calculated in good physical approximation. With these settings and an NVIDIA RTX 4090 GPU, render times per image were approximately 10 s without the water block and 50 s including it. The longer runtime is explained by the more complex light paths in the volumetric medium. This demonstrates the software's efficient implementation, as it must compute up to 3.2 billion primary rays per image at a resolution of 1024×768 with a maximum of 4096 samples per pixel.

Figure 3. Side view in the Blender scene. From this underwater viewpoint, total refraction at the water surface can be seen, highlighting the ray tracing capabilities of Blender. In top of the background some camera positions are visible.

4. Scene generation

In this Section, we describe the point cloud generation using MVS (see Section 4.1) and NeRFs (see Section 4.2).

4.1 MVS point cloud generation

In order to compute a dense 3D point cloud from the rendered images, the SfM workflow was first executed in Agisoft Metashape to extract tie points that serve as the initialization for the subsequent MVS processing. Since the true values of the cameras' exterior orientation parameters are known a priori, their respective weights in the bundle adjustment were set to very high values to prevent any deviation. As a result, the estimated camera positions deviate by no more than 1 μm in the no-water scene and 0.1 mm in the underwater scene from the imported reference values. Likewise, the usual in-situ calibration, where the focal length, principal point, and distortion parameters are refined, was omitted, because no optical aberrations were simulated and the nominal camera model is to be propagated unchanged. Although future investigations could incorporate lens aberrations (e.g., chromatic and spherical aberration, sensor curvature, and various distortion effects (CANON INC., 2006)) in Blender to further enhance realism, we deliberately opted to exclude these effects in order to isolate the influence of water refraction. Upon completion of the SfM step, dense point clouds were generated via MVS algorithms. However, due to the limited number of low-resolution images, only about six million points were reconstructed in both scenarios. Furthermore, because the area of interest was cropped to its central region, we also exported points derived from as few as two overlapping depth maps to maximize detail at the expense of an increased risk of erroneous 3D coordinate estimates.

4.2 NeRF point cloud generation

Since underwater scenes frequently violate the photogrammetric assumptions required for robust MVS reconstruction, the datasets were additionally processed using a NeRF-based approach, which has been shown to yield superior results in such environments (Remondino et al., 2023). NeRFs are particularly beneficial for capturing sub-pixel structures and dynamic elements, such as moving vegetation or water surfaces, during data acquisition. To process and train the NeRFs, we employed the modular PyTorch framework Nerfstudio (Tancik et al., 2023), which natively imports the camera poses and intrinsic parameters exported from Agisoft Metashape after bundle adjustment. This ensures that both MVS and NeRF reconstructions share identical georeferencing and camera calibration, facilitating direct comparison. Nerfstudio further provides modular NeRF components, a real-time WebGL viewer for visual

inspection, and a built-in point cloud export utility. For network training, we selected the default Nerfacto model, which 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) and simultaneously performs a pose refinement of all camera positions. Both the model without and with a water body were trained for 30000 iterations. After this, point clouds were exported in the global (world) coordinate frame with 50 million points and a statistical outlier removal (SOR) threshold of 2.0.

We evaluated view synthesis quality on water and no-water datasets using three image-based metrics: PSNR for pixel accuracy, SSIM for structural similarity, and LPIPS for perceptual consistency. Each synthesized NeRF view was compared to its corresponding Blender ground-truth image.

5. Evaluation

To validate the reconstruction accuracy of the point clouds, we employed a C2M comparison implemented in CloudCompare, using the original combined reference mesh of the scene against the four derived point clouds (MVS with and without water and NeRF with and without water). For each point of the reconstructed point cloud, C2M computes the closest distance to a triangle in the reference mesh. All datasets, including the reference mesh and the exported point clouds, were clipped to the core area ($75 \times 75\ m^2$). The NeRF point clouds had initially been exported with 50 million points. To enable a fair comparison with MVS, they were subsampled to a ground sampling distance of 5 cm for the no-water scene and 6.5 cm for the water scene, resulting in approximately 2.5 million points for all four point clouds. After subsampling, a noise filter was applied to the NeRF-derived point cloud containing water. This filter operates as a low-pass filter by fitting a local plane around each point and removing points that lie beyond a relative distance of $2\ \sigma$ from the plane. Although the point cloud was already filtered with an SOR-filter in the Nerfstudio export, the noise filter further improves the quality of the point cloud. This additional filtering was not necessary for the other three point clouds. Additionally, the NeRF point clouds required a vertical (z-axis) offset of +18.9 cm, a shift that has been observed in prior investigations (Winiwarter et al., 2025) and whose origin, potentially stemming from additional pose refinement in Nerfstudio or differences in the interpretation of the camera parameters between Nerfstudio and Metashape, remains an open question. Unlike most studies in this field, no further registration (e.g. using ICP (Besl and McKay, 1992)) was necessary, since all point clouds share the same coordinate reference system.

The C2M distances serve as the foundation for the completeness metric. Building on the framework of Seitz et al. (2006), completeness was defined as the proportion of ground-truth points G that fall within a user-specified tolerance d of the reconstructed surface R. The threshold d is selected to encompass acceptable reconstruction errors, ensuring that noisier reconstructions naturally yield lower completeness scores. For d we used the value of 0.2 m suggested by Hermann et al. (2024) for aerial scenes, which in our case lies significantly above the GSD of the point clouds. Because completeness should be primarily evaluated for the ship, the reference mesh was cropped to its central $25 \times 30\ m^2$ region and uniformly sampled with 15 million surface points G_p to ensure a substantially higher point count than in any reconstructed cloud R. For each of the four R_i point clouds, we then computed the Cloud-to-Cloud distances for every reference point in G_{pi} to its nearest neighbor in R_i, discarded all points whose distance exceeded the tolerance d, and defined completeness as the ratio of the remaining points $G_{p<d}$ to the total number of reference points G_p.

6. Results and discussion

In the following, the results of the scenes with water and without water are presented. In Section 6.1, the results of the 2D analyses are presented; Section 6.2 details the outcomes of the 3D evaluation with C2M; and Section 6.3 reports the findings of the completeness evaluation.

6.1 2D evaluation metrics

The results of the 2D image evaluation are summarized in Table 3. Contrary to expectations, both PSNR and SSIM are higher for the water dataset than for the no-water dataset, whereas only the LPIPS score is better for the no-water case. These findings suggest that, despite refraction effects, the water images yielded ostensibly superior training performance. Accordingly, caution is warranted when solely using these image-based metrics in a geodetic context, as they appear to correlate poorly with the reconstruction quality of the underlying geometry. However, the relevance of this conclusion in real-world conditions is limited, since natural water bodies never exhibit perfectly smooth, particle-free surfaces, and because Nerfacto's pose-refinement step may partially absorb or compensate for refraction effects. A more reliable assessment is therefore provided by the C2M comparison presented in the next section.

Dataset	PSNR ↑	SSIM ↑	LPIPS ↓
No-water	23.34	0.52	0.11
Water	24.52	0.53	0.17

Table 3. Results of view-synthesis quality metrics for no-water and water datasets.

6.2 3D evaluation metrics

To quantify the overall reconstruction accuracy of the four point clouds, C2M distances are calculated. Table 4 reports the mean and standard deviation (Std) of the C2M distances after excluding the outermost 10% of samples to prevent a high influence of outliers. In the no-water scene, both methods achieve excellent agreement with the reference mesh. As shown in Table 4, the MVS point cloud attains a mean deviation of less than 1 mm and a standard deviation below 1 cm, well under the planned GSD of 3 cm, while the NeRF point cloud, after applying the vertical shift, yields a comparable mean error (3 cm) but a larger scatter (Std of 6 cm). Expected accuracies are typically 0.5 - 1 pixel (1.5 - 3 cm) in position and 2 - 3 pixels (6 - 9 cm) in height for nadir images (Luhmann, 2023). In both cases, the point clouds lie within these accuracy thresholds, and the distance-histograms closely follow a Gaussian profile with a mean close to zero, indicating that residual errors are dominated by random noise rather than systematic effects.

In contrast, in the submerged water scenario, the variability of the C2M distances increases markedly. Both MVS and NeRF

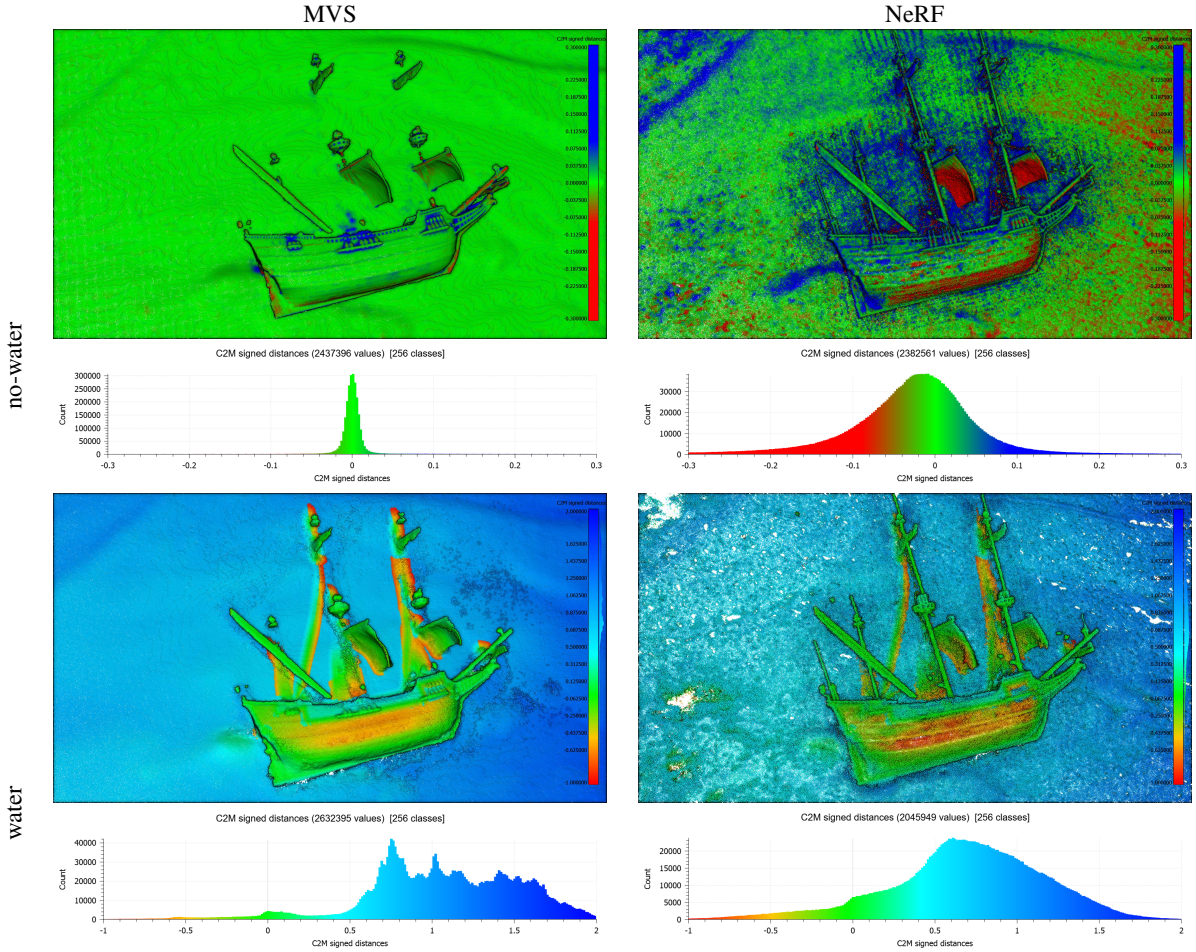

Figure 4. Results of the C2M distances. The results of both methods for the scene without water are shown at the top and the results of the scene with water are shown at the bottom. The histogram with the distances [m] is shown below each of the point clouds. The color scales match in the histogram and point cloud, but differ between the two scenes.

exhibit standard deviations in the order of 35 cm, and their mean deviations rise to approximately 1.1 m for MVS and 0.7 m for NeRF. Here, the distribution functions deviate significantly from normality, pointing to systematic effects during reconstruction. A closer inspection of Figure 4 reveals that regions of the ship closest to the water surface exhibit smaller discrepancies. This is consistent with Snell's law of refraction, whereby light rays bend toward the normal upon entering water. As depth increases, the angular compression induced by refraction causes triangulated points to be displaced deeper than their true positions, producing a depth-dependent bias.

In summary, at a water depth of 3.5 m, both reconstruction workflows result in a systematic refractive bias of roughly 1 m, which must be corrected to yield usable 3D data. Furthermore, while MVS outperforms NeRF in the no-water experiment, NeRF demonstrates a more Gaussian error distribution underwater, suggesting its greater robustness to refraction-induced distortions.

6.3 Completeness

As shown in Table 4, a consistent trend in completeness emerges across both scenarios. In the no-water scene, MVS achieves 15% higher completeness than in the water scene, whereas NeRF exhibits a 5% difference. Moreover, NeRF point clouds are substantially more complete than those from MVS, by 9% in no-water and by 21% in water scene. This disparity is particularly evident in Figure 5. MVS fails to reconstruct the masts and large portions of the ship's deck in both environments, while NeRF successfully recovers the entire mast in the no-water dataset and most of it in the water dataset. The deck is accurately represented in both NeRF-cases, with only the thin ropes of the mast remaining unresolved, which is not surprising, as as their thickness falls short of the planned GSD. In summary, these results demonstrate that refraction effects and, more importantly, the reconstruction method have a significant impact on geometric fidelity. NeRF-based reconstruction shows a clear advantage in this simulated, aerial data.

Dataset	Mean [m]	Std [m]	Compl. [%]
MVS: No-water	0.0003	0.006	79.1
NeRF: No-water	*-0.030*	0.058	88.1
MVS: Water	1.092	0.346	63.8
NeRF: Water	*0.706*	0.360	83.1

Table 4. Results of the mean and the standard deviation (Std) [1 σ] of the C2M distances and the completeness evaluation for no-water and water datasets. The mean values of the NeRF datasets are not representative, as they were previously manually shifted in height. For estimating the completeness, a distance threshold of 0.2 m was used.

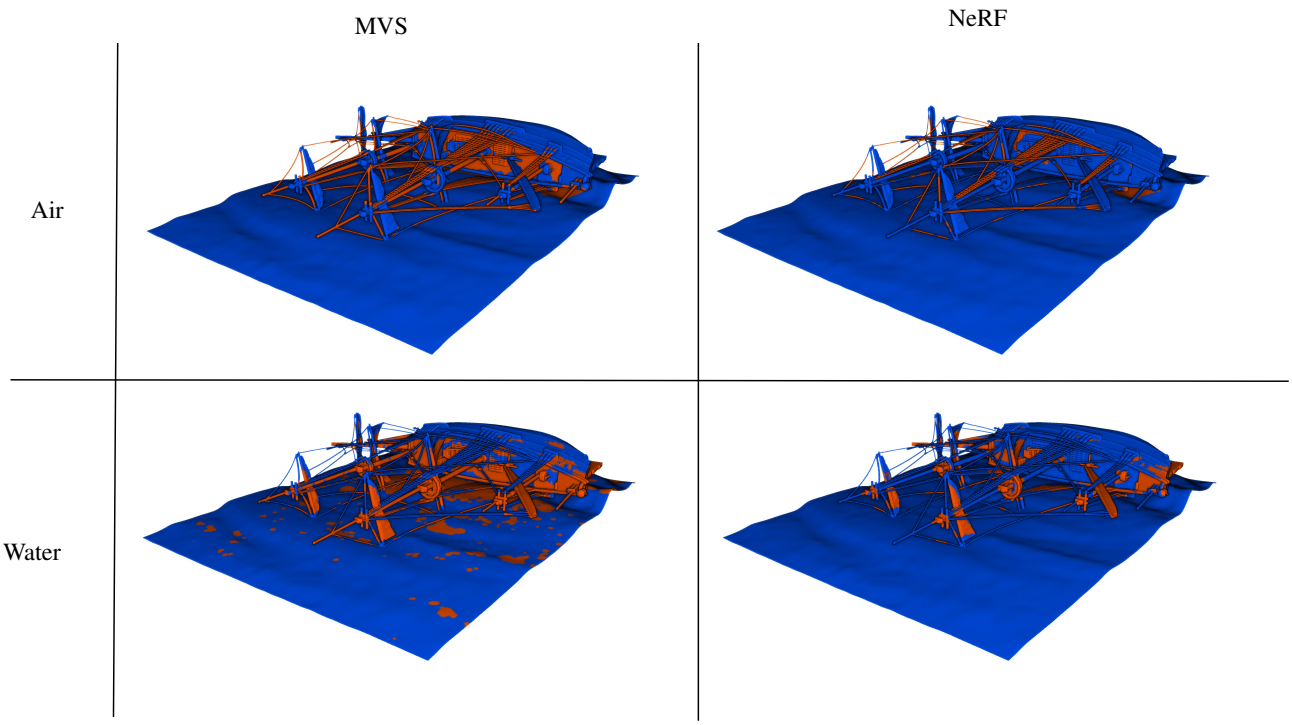

Figure 5. Results of the completeness evaluation. The results of both methods for the scene without water are shown at the top and the results of the scene with water are shown at the bottom. Blue indicates the regions that could be reconstructed within the threshold and red shows regions that could not be reconstructed from the images.

7. Conclusion and Outlook

This study developed a framework to simulate 3D underwater scenes. Camera positions were computed in Python and used in Blender to render images. These were processed with MVS and NeRF to generate and evaluate dense point clouds for geometric accuracy and completeness. The results demonstrate that refraction at the air-water surface induces a systematic depth bias of approximately 1 m at 3.5 m water depth, underscoring the necessity of refraction correction for reliable underwater reconstruction. While MVS outperforms NeRF in the absence of water, NeRF exhibit superior geometric accuracy and substantially higher completeness in submerged conditions.

To further bridge the gap between synthetic experiments and real-world UAV surveys, future work in the BathyNeRF research project will focus on enhancing the realism of the simulation. Integration both static and dynamic water surfaces caused by waves, modeling detailed camera and lens optical properties and implementing physically based lighting with reflections at the water surface are the next steps. This will provide a more realistic base for developing and benchmarking novel underwater photogrammetric algorithms. Another future approach would involve fusing above-water MVS, derived point clouds, with below-water NeRF generated reconstructions.

Acknowledgments

The research project BathyNeRF is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 538522540 and by the Austrian Science Fund (FWF) 10.55776/ PIN1353223. It is a collaboration of the Karlsruhe Institute of Technology, TU Wien, and University of Innsbruck.

References

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.

Bartell, F. O., Dereniak, E. L., Wolfe, W. L., 1981. The Theory And Measurement Of Bidirectional Reflectance Distribution Function (Brdf) And Bidirectional Transmittance Distribution Function (BTDF). *SPIE 0257, Radiation Scattering in Optical Systems*. https://doi.org/10.1117/12.959611.

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.

Blender Development Team, 2024. Blender. version: 4.3.2. https://www.blender.org (26.05.2025).

Brezovsky, M., Guenthner, A., Schulte, F., Winiwarter, L., Jutzi, B., Mandlburger, G., 2025. Analysis of refraction-aware neural radiance fields for 3D reconstruction of underwater scenes. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*.

CANON INC., 2006. *EF LENS WORK III - Die Augen von EOS*. 8. Auflage.

Chaitanya, C. R. A., Kaplanyan, A. S., Schied, C., Salvi, M., Lefohn, A., Nowrouzezahrai, D., Aila, T., 2017. Interactive reconstruction of Monte Carlo image sequences using a recurrent denoising autoencoder. *ACM Transactions on Graphics*, 36(4), 1-12. https://doi.org/10.1145/3072959.3073601.

DJI Enterprise, 2021. Aerial surveying just got smart. https://enterprise-insights.dji.com/blog/smart-oblique-capture (26.05.2025).

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/10.1561/0600000052.

Guenthner, A., Brezovsky, M., Schulte, F., Winiwarter, L., Mandlburger, G., Jutzi, B., 2025. Exploring the Potential of NeRFrac for Photogrammetric Bathymetry: First Application to UAV-based Data from the Pielach River. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*.

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.

Lind, L., Eckstein, R. L., Relyea, R. A., 2022. Direct and indirect effects of climate change on distribution and community composition of macrophytes in lentic systems. *Biological Reviews*, 97(4), 1677–1690. https://doi.org/10.1111/brv.12858.

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.

Luhmann, T., 2023. *Nahbereichsphotogrammetrie: Grundlagen - Methoden - Beispiele*. 5 edn, Wichmann.

Mandlburger, G., Rhomberg-Kauert, J., Gueguen, L.-A., Mulsow, C., Brezovsky, M., Dammert, L., Haines, J., Glas, S., Schulte, F., Amon, P., Winiwarter, L., Jutzi, B., Maas, H.-G., 2025. Mapping shallow inland running waters with UAV-borne photo and laser bathymetry. *Journal of Applied Hydrography – Hydrographische Nachrichten*, 130, 42–53. https://doi.org/10.23784/HN130-06.

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.

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.

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.

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.

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$_3$1.

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.

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.

Triggs, B., McLauchlan, P. F., Hartley, R. I., Fitzgibbon, A. W., 2000. Bundle adjustment — a modern synthesis. G. Goos, J. Hartmanis, J. Van Leeuwen (eds), *Vision Algorithms: Theory and Practice*, 1883, Springer Berlin Heidelberg, 298–372.

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.

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.

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.

Zhan, Y., Nobuhara, S., Nishino, K., Zheng, Y., 2023. NeRFrac: Neural Radiance Fields through Refractive Surface. *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 18356-18366. https://doi.org/10.1109/ICCV51070.2023.01687.