Highlights

BathyFacto: Refraction-Aware Two-Media Neural Radiance Fields for Bathymetry

Markus Brezovsky, Anatol Günthner, Frederik Schulte, Lukas Winiwarter, Boris Jutzi, Gottfried Mandlburger

- Refraction-aware two-media NeRF for photogrammetric bathymetry.
- Explicit image ray tracing through an open water surface and Snell's-law refraction correction.
- Single-sampler architecture with kinked density for unified air–water sample allocation.
- Refraction-corrected point cloud export integrated into Nerfstudio for metric bathymetry.

BathyFacto: Refraction-Aware Two-Media Neural Radiance Fields for Bathymetry

Markus Brezovsky[a,*], Anatol Günthner[b], Frederik Schulte[c], Lukas Winiwarter[c], Boris Jutzi[b] and Gottfried Mandlburger[a]

[a]Department of Geodesy and Geoinformation, TU Wien, Vienna, Austria
[b]Institute of Photogrammetry and Remote Sensing (IPF), Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
[c]Unit of Geometry and Surveying, University of Innsbruck, Innsbruck, Austria

ARTICLE INFO

Keywords:
Neural Radiance Fields
Refraction
Multimedia Photogrammetry
Two-Media Rendering
Point Cloud
Bathymetry

ABSTRACT

Through-water photogrammetry based on UAV imagery enables shallow-water bathymetry, but refraction at the air-water interface violates the straight-ray assumption of Structure-from-Motion and causes systematic depth bias. We present BathyFacto, a refraction-aware two-media extension of Nerfacto integrated into Nerfstudio that targets metrically precise underwater point clouds. BathyFacto uses a shared hash-grid-based density field with a medium-conditioned color head that receives a one-bit medium flag (air or water) and traces each camera ray as two segments: a straight segment in air up to a planar water surface and a refracted segment in water computed via Snell's law with known refractive indices. To allocate samples efficiently across the air–water boundary, we employ a single proposal-network sampler that operates on a virtual straight ray spanning both media, combined with a kinked density wrapper that transparently corrects water-segment positions along the refracted direction before density evaluation. A data adaptation pipeline converts photogrammetric reconstructions to a Nerfstudio-compatible format, estimates the water plane from boundary markers, and provides per-pixel medium masks to gate refraction. We also extend the point cloud export with refraction-corrected backprojection and reversible coordinate transforms to world and global frames. On a simulated two-media scene with known ground truth, BathyFacto with refraction achieves a Cloud-to-Mesh mean distance of 0.06 m and 87 % completeness, compared to 0.52 m / 29 % for the Nerfacto baseline and 0.36 m / 21 % for conventional Multi-View-Stereo (MVS) without refraction correction.

1. Introduction

Shallow-water bathymetry is essential for, among others, hydrodynamic-numerical modeling, flood risk assessment, habitat mapping, and infrastructure planning along rivers, lakes, and coastal zones. Unoccupied aerial vehicles (UAVs) equipped with consumer-grade cameras offer a cost-effective and flexible platform to acquire high-resolution imagery of such environments (Westoby et al., 2012). Structure-from-Motion (SfM) and Multi-View Stereo (MVS) photogrammetry can then recover dense 3D point clouds from these images, enabling rapid survey workflows (Schönberger and Frahm, 2016; Hirschmuller, 2008). This has also been explored for shallow-water mapping when imaging conditions are suitable (Dietrich, 2017; Del Savio et al., 2023).

However, through-water photogrammetry faces a fundamental challenge: light rays refract at the air-water interface according to Snell's law, bending toward the surface normal as they enter the denser medium. Standard SfM-MVS pipelines assume straight ray paths and therefore produce systematically biased underwater geometry reconstructions in multimedia settings (Maas, 2015). Although post-processing corrections based on known water surface geometry and viewing angles can reduce this bias (Agrafiotis et al., 2020; Lingua et al., 2023), they require handling of each image ray and remain sensitive to uncertainties in surface estimation.

Active sensing with airborne laser bathymetry (ALB) provides an alternative reference by measuring range through water using a pulsed green laser that delivers depth measurements in suitable conditions. However, it is expensive, requires specialized sensors and is limited by legal, environmental and surface conditions (Mandlburger, 2022). Passive, image-based methods therefore remain attractive for routine monitoring, provided refraction can be modeled adequately.

Neural Radiance Fields (NeRFs) enable novel-view synthesis by representing scenes as continuous volumetric functions learned from posed images (Mildenhall et al., 2020). Extensions such as Mip-NeRF 360 (Barron et al., 2022) handle unbounded scenes and anti-aliasing, while Nerfstudio (Tancik et al., 2023) provides a modular framework that supports research and deployment. Complementary advances in 3D Gaussian Splatting (Kerbl et al., 2023) achieve real-time rendering, and underwater-focused methods like SeaThru-NeRF (Levy et al., 2023) address scattering and color restoration. However, all of these NeRF variants assume a single homogeneous medium and do not model refraction at interfaces.

NeRFrac (Zhan et al., 2023) explicitly incorporates a refractive surface into the NeRF formulation. It learns both the scene radiance field and the distance to a potentially wavy, refractive interface, enabling novel-view synthesis of underwater scenes observed through water. While NeRFrac reports qualitative results, it focuses on rendering quality rather than geometric accuracy and does not provide a

*Corresponding author
markus.brezovsky@tuwien.ac.at (M. Brezovsky)
ORCID(s):

pipeline for exporting metrically meaningful point clouds suitable for bathymetric applications. Moreover, its Vanilla-NeRF backbone does not incorporate modern sampling strategies such as proposal networks, which can limit convergence speed and stability on datasets with large viewpoint variation.

In our contribution, we present **BathyFacto**, a refraction-aware two-media extension of Nerfacto specifically designed for photogrammetric bathymetry. BathyFacto is integrated into Nerfstudio and targets improved geometric accuracy of underwater point clouds with refraction-corrected export in reproducible coordinate frames. Our main contributions are as follows:

Two-media Nerfacto with ray refraction at the air-water interface based on Snell's law. We extend Nerfacto with a shared hash-grid-based density field and a medium-conditioned color head that receives a one-bit medium flag (air or water). Each camera ray is traced as two piecewise linear segments, first in air and a second in water, refracted at the air-water interface via Snell's law with known refractive indices.

Single-sampler architecture with kinked density. Rather than maintaining separate sampling hierarchies for air and water, we employ a single proposal-network sampler that operates on a virtual straight ray spanning both media. A kinked density wrapper transparently corrects water-segment positions along the refracted direction before density evaluation, enabling the proposal network to adaptively allocate samples across the air–water boundary.

Nerfstudio integration. We implement BathyFacto within the Nerfstudio framework, leveraging its hash-grid encoding, proposal-network sampling, camera pose optimization, and training infrastructure. The accompanying data pipeline supports per-pixel medium masks for mask-gated two-media rendering and export. This integration promotes reproducibility and allows practitioners to benefit from ongoing Nerfstudio improvements.

Refraction-corrected point cloud export. We extend the standard NeRF point cloud exporter to back-project depth predictions along the refracted two-segment ray geometry. Reversible coordinate transformations allow for exporting in a normalized global reference frame for direct comparison with reference geometry (e.g., simulation ground truth).

Ablation study. We evaluate BathyFacto on a simulated two-media scene with known ground truth, reporting 3D point cloud metrics. We compare configurations with and without refraction correction to quantify the benefit of explicit two-media modeling.

The remainder of this contribution is organized as follows. Section 2 reviews related work on through-water photogrammetry, standard NeRFs, and refractive NeRFs. Section 3 describes the BathyFacto methodology, starting with the adaptation of the data and continuing to the shared-geometry two-media architecture, interface coupling, and refraction-corrected point cloud export. Section 4 details the simulation dataset providing ground truth, describes the hyperparameters and model variants (Nerfacto baseline, BathyFacto with and without refraction), and outlines the evaluation metrics. The results derived by BathyFacto are presented in Section 5. In Section 6 our proposed method is discussed and the practical strengths and limitations are summarized. With Section 7 the contribution concludes and an outlook to future research is provided.

2. Related Work

This section situates BathyFacto within prior work on optical bathymetry and through-water photogrammetry, standard NeRF reconstruction, and NeRF variants that explicitly model refraction. We first summarize established correction strategies and sensing modalities in multimedia settings, then review NeRF foundations and recent refractive extensions to motivate the need for explicit two-media ray modeling in reconstructions.

2.1. Through-Water Photogrammetry

UAV-based photogrammetry provides a flexible and cost-effective option for high-resolution topographic mapping (Westoby et al., 2012). Whenever image rays traverse more than one medium with different refractive indices, the interface violates the straight-ray assumption underlying standard SfM and MVS, because rays refract according to Snell's law as they enter the denser medium (Del Savio et al., 2023). In photogrammetric terms, the collinearity condition is not met in multimedia settings, and ignoring refraction introduces systematic geometric errors (Maas, 2015).

Passive correction strategies typically estimate the water surface geometry and then correct underwater rays either as a post-processing step or within the optimization. Image-/ray-space correction approaches reproject underwater observations using a planar or locally estimated interface and camera geometry (Agrafiotis et al., 2020). Iterative refraction-correction workflows apply Snell's law together with camera viewing geometry to adjust SfM-MVS-derived bathymetric point clouds and reduce depth bias (Lingua et al., 2023). In practice, the achievable accuracy depends on the validity of the assumed interface model and the quality of the image texture in water (Maas, 2015; Agrafiotis et al., 2020), which is often hampered by sediment-induced turbidity.

Active sensing provides an alternative reference standard. Airborne laser bathymetry (ALB) directly measures range through water using a pulsed green laser and post-processing properly accounts for two-media propagation effects (Guenther et al., 2000; Mandlburger, 2022). Although ALB provides depth measurements under suitable conditions, it requires specialized sensors and flight operations. For UAV-based image surveys, a remaining challenge is to couple a physically correct two-media image formation model with 3D reconstruction pipelines such that the exported underwater geometry is metrically interpretable.

2.2. Photogrammetric Workflow

The estimation of intrinsic and extrinsic camera orientations is a fundamental prerequisite for downstream 3D reconstruction frameworks, such as MVS and NeRFs. This requirement is traditionally addressed through SfM, a photogrammetric workflow that reconstructs 3D scenes from 2D imagery by simultaneously estimating camera poses and a sparse tie point cloud (Schönberger and Frahm, 2016). The process is initiated by extracting unique image features, typically via the Scale Invariant Feature Transform (SIFT) algorithm (Lowe, 2004). Subsequently, these descriptors are matched in stereoscopic image pairs using robust estimation techniques, such as Random Sample Consensus (RANSAC) (Fischler and Bolles, 1981), to filter out outliers. Following the derivation of relative camera poses from these matches, the pipeline culminates in a global bundle adjustment (Triggs et al., 2000). This optimization refines the interior and exterior camera parameters and 3D point locations simultaneously by minimizing the total reprojection error, thereby ensuring the geometric fidelity of the calibration and the resulting point cloud. MVS constitutes a framework that builds upon the SfM workflow to derive a dense point cloud by leveraging sparse tie points in conjunction with the intrinsic and extrinsic camera orientation parameters (Schönberger et al., 2016; Hirschmuller, 2008). The MVS process starts with the detection of corresponding image points constrained by epipolar geometry, subsequently transforming these correspondences into disparity maps which are then converted into per-pixel depth information using the known camera parameters. To generate a comprehensive and dense 3D representation of the scene, these individual depth maps are fused and integrated (Furukawa and Hernández, 2015). On well-textured surfaces, these algorithms exhibit highly accurate surface reconstructions. Conversely, the fidelity of the reconstruction decreases drastically or may fail entirely under suboptimal illumination, particularly on textureless or refractive surfaces, such as those encountered in bathymetric scenarios.

2.3. Neural Radiance Fields

NeRFs model scenes as continuous volumetric functions that map 3D position and viewing direction to density and radiance, enabling high-quality novel-view synthesis from oriented images (Mildenhall et al., 2020). For large-scale captures, extensions such as Mip-NeRF 360 (Barron et al., 2022) improve sampling and stability in unbounded scenes, and widely used frameworks such as Nerfstudio with its Nerfacto variant standardize training, evaluation, and deployment (Tancik et al., 2023). In parallel, 3D Gaussian Splatting provides a complementary point-based radiance field representation with real-time rendering performance (Kerbl et al., 2023). These developments have improved reconstruction quality and usability; however, most NeRF variants still assume straight rays in a single homogeneous medium.

For photogrammetric bathymetry, this assumption is problematic: the refractive interface is often a dominant error source, not only the representational capacity of the radiance field. Consequently, the key requirement is not merely plausible view synthesis but a reconstruction pipeline that maintains geometric fidelity under two-media ray bending and supports metric 3D geometry export, both for comparison to external references and for applications that explicitly target underwater geometry rather than only image quality. In the following, we therefore review refraction-aware NeRF approaches.

2.4. Refractive NeRFs

Recent NeRF literature addresses non-Lambertian effects and refractive phenomena from different perspectives. Ref-NeRF improves view synthesis of glossy surfaces by explicitly parameterizing view-dependent radiance via reflected directions (Verbin et al., 2022), and REF^2-NeRF extends this direction to jointly model reflection and refraction in NeRF-style rendering (Kim et al., 2023). While these models improve appearance under complex light transport, they primarily target photorealistic rendering and often rely on implicit or simplified geometric mechanisms rather than physically-grounded explicit two-media ray tracing.

More directly related to BathyFacto are approaches that incorporate ray bending. LB-NeRF addresses light bending in transparent media within a NeRF-style formulation by learning an offset field that displaces sampling points away from the straight camera ray (Fujitomi et al., 2022). This implicitly represents refraction without directly enforcing Snell's law or an explicit interface geometry: a separate network predicts per-sample 3D offsets as a function of position and view direction, regularized to limit ambiguity in the learned offsets. While this makes the method lightweight and independent of prior knowledge on refractive indices or surface shape, it also decouples ray bending from a physically constrained two-media model and is demonstrated on object-centric scenes with compact transparent volumes rather than extended air-water interfaces typical of UAV bathymetry. NeRFrac explicitly embeds Snell's law into the NeRF pipeline by modeling a refractive interface of planar or smoothly varying shape and tracing rays as two segments, with joint optimization of interface shape and radiance field (Zhan et al., 2023). The method is evaluated on synthetic underwater scenes generated by ray tracing and on real data captured with a fixed multi-camera array above a water tank, showing that physically based refraction modeling improves novel-view synthesis and enables reconstruction of the water surface itself. However, the focus remains on image-space quality. The approach is not designed for point cloud export and has not been evaluated on large-scale UAV imagery.

A subsequent extension refines this method by improving refractive surface reconstruction and rendering quality through ray-wise Bezier surface fitting of the interface, adaptive hierarchical sampling of refracted rays, and a hybrid directional encoding based on spherical harmonics (Zhang et al., 2026). The evaluation nevertheless remains restricted to small-scale synthetic and laboratory water tank setups

without point cloud export, and thus does not address 3D scene reconstruction from real-world UAV imagery.

For UAV-based bathymetry, NeRFrac has first been applied to outdoor river scenes by training on nadir UAV imagery over a pre-Alpine River and evaluating its refractive modeling exclusively with 2D image-space metrics, without exporting or validating 3D point clouds (Günthner et al., 2025). Subsequent work extends NeRFrac with a dedicated point cloud export and qualitatively inspects refractive 3D reconstruction from UAV data, but the resulting geometry remains in a normalized internal coordinate system and has not yet been metrically validated against external references (Brezovsky et al., 2025).

In summary, existing refractive NeRFs demonstrate the benefits of explicit refraction modeling on synthetic and laboratory data, but they either rely on implicit offset fields without physical constraints or are tailored to controlled camera arrays and do not yet provide a Nerfstudio-integrated workflow for metrically interpretable 3D reconstruction from UAV imagery. BathyFacto addresses this gap by extending the Nerfacto backbone with a physically constrained two-media ray model, synthetic datasets that emulate UAV acquisition geometries, and a refraction-aware point cloud export tailored to photogrammetric bathymetry.

3. Methodology

This section describes the BathyFacto pipeline which is depicted in Figure 1. Our workflow consists of three stages: i) converting photogrammetric outputs into a Nerfstudio-compatible dataset, including water-surface metadata, as detailed in Section 3.1; ii) training BathyFacto as a two-media Nerfacto extension that renders each camera ray with a straight segment in air and, where applicable, a refracted segment in water, described in Sections 3.2-3.5; and iii) exporting a refraction-corrected point cloud via two-segment backprojection in reversible coordinate frames, outlined in Section 3.6.

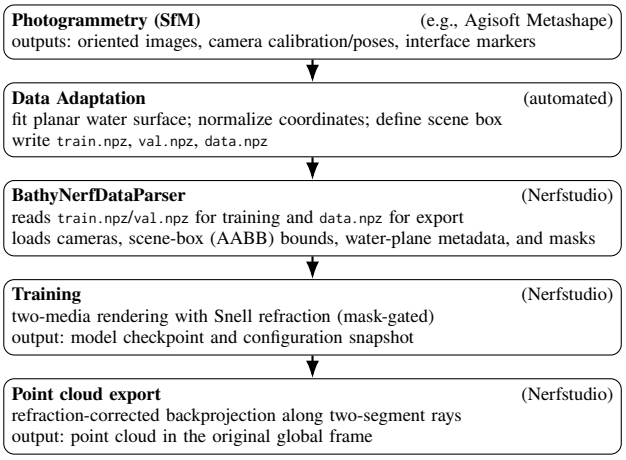

Figure 1: BathyFacto workflow from photogrammetric reconstruction to refraction-corrected point cloud export. Photogrammetry outputs are converted into NumPy archives (`.npz`). The BathyNerfDataParser loads `train.npz`/`val.npz` for training and `data.npz` for full-scene export. Finally, the refraction-aware exporter back-projects points along the two-segment ray geometry and maps them back to the original global frame.

3.1. Data Adaptation

BathyFacto operates on Nerfstudio-style datasets stored as NumPy archives (`.npz`-files) containing cameras, scene bounds, and auxiliary metadata. Starting from a photogrammetric reconstruction with oriented images, we perform the following steps to obtain these files:

Photogrammetric Export. We export reconstructed camera intrinsic and extrinsic parameters, as well as 3D markers representing the water surface, which are, for now, placed manually along the land-water boundary. We use Agisoft Metashape and export it as an `.xml`-file.

Pose Transformation. The camera poses and 3D marker points are converted from the photogrammetric coordinate system into the OpenCV-style convention used by Nerfstudio.

Water Surface Estimation. Marker points along the shoreline are used to fit a planar water surface via least-squares adjustment, yielding the plane normal vector and intercept in the photogrammetric frame.

Scene Transformation. The cameras and markers are first rotated so that the fitted water plane becomes horizontal and the positive z-axis points upward towards the cameras. We then translate the scene, shifting the coordinate origin to the centroid of the camera positions, and apply a single global scale factor so that all camera and marker coordinates lie approximately within $[-1, 1]$, resulting in numerically stable units in the training space.

Scene Box. The scene box used for NeRF sampling is derived from the horizontal (x, y) extent of the normalized camera positions and the vertical (z) range spanned by both cameras and markers. This axis-aligned box is then translated to move its center into the markers' centroid, ensuring that sampling covers both the air and underwater regions in a

balanced way. The camera positions, the water surface plane, and the resulting scene box are visualized in Figure 2.

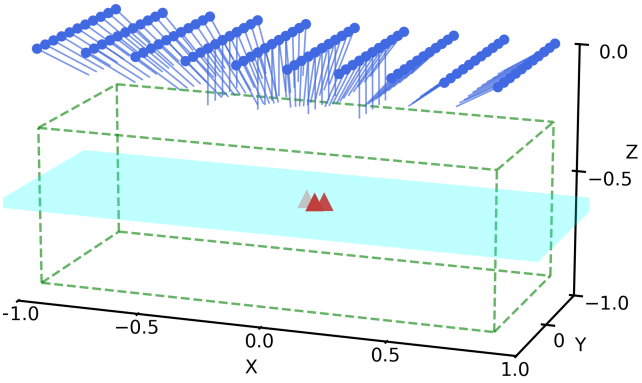

Figure 2: Acquisition geometry. Camera positions and viewing directions are shown in blue. Markers (red) define the water surface (cyan), from which the scene box (dashed green) is derived.

Medium Masks For mixed land-water scenes, BathyFacto can optionally use per-image masks to restrict refractive rendering to water-covered pixels while keeping land pixels on straight, non-refracted rays. In our synthetic dataset, such masks are available. For other datasets, water masks must be generated in a preprocessing step. Each mask file must share the same filename as its corresponding RGB image, and must be aligned pixel-wise to ensure consistent gating of refractive vs. non-refractive rays during training and export. An example water mask for one image is shown in Figure 3.

Train/Validation Split and Packaging. In a final step, the oriented images are split into a training subset and a validation subset, with nadir and oblique views distributed as evenly as possible across both sets. We use 90% of all images for training and the remaining images for validation. For each split, the corresponding RGB filenames, medium masks, camera intrinsics, and extrinsics are written to `train.npz` and `val.npz`, respectively. In addition, a full `data.npz` file stores all images, all masks, the complete camera block, and the shared scene metadata, including the water surface parameters, the normalization transform, and the scene box.

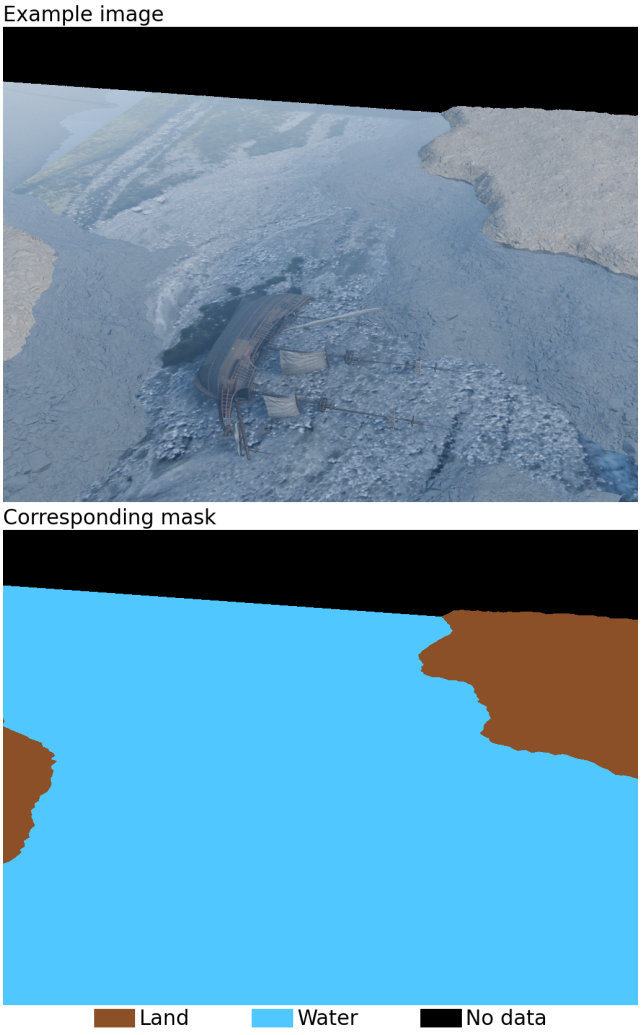

Figure 3: Example image and corresponding mask used to gate two-media rendering. The mask assigns each pixel to land (air-only), water (two-segment ray with refraction), or ignore/no-sampling regions that are excluded from training losses and evaluation.

3.2. BathyFacto Architecture

BathyFacto extends Nerfacto with a shared-geometry two-media field that uses a single density representation for both air and water, combined with a medium-conditioned color head. This design enforces geometric consistency across the air-water interface: the same hash-grid-based density field predicts volume density σ everywhere, while a single color MLP receives a one-bit medium flag (0 for air, 1 for water) as additional input to decode medium-specific appearance. Given a sample position $\mathbf{x}$, direction $\mathbf{d}$, and medium flag $m \in \{0, 1\}$, the shared field predicts density σ and a geometry feature embedding $\mathbf{h}$, which is then passed along with m to the color head to produce a medium-aware RGB color $\mathbf{c}$.

We sample points along rays within the scene box. BathyFacto uses Nerfacto-style proposal-network sampling to focus samples around high-density regions. An Axis Aligned

Bounding Box (AABB) near/far collider provides per-ray bounds.

3.2.1. Shared Field, Encodings, and Physics Parameters

The main field uses a multi-resolution hash-grid encoding (Müller et al., 2022) with 16 resolution levels, a base resolution of 16, a maximum resolution of 2048, and two features per level, stored in a hash table with a size of 2^{19}. The density MLP has a hidden dimension of 64. This encoding replaces the sinusoidal positional encoding used in vanilla NeRF and provides substantially faster convergence and higher spatial fidelity.

The shared density MLP produces both volume density and a geometry feature embedding. A single color head takes the geometry embedding, an encoded viewing direction, optional appearance embeddings, and a one-bit medium flag as input and predicts RGB values. The medium flag (0 for air, 1 for water) is concatenated to the color head input, increasing its input dimension by one. This lightweight conditioning enables the color head to learn medium-specific appearance (such as water absorption and color shift) while sharing all weights across both media, which is more parameter-efficient than maintaining two independent color heads. Optional per-image appearance embeddings (dimension 32) allow BathyFacto to accommodate photometric variation across the image set.

The two-media physics is controlled by fixed refractive indices n_{air} and n_{water} (default: 1.0 and 1.333). In the ablation used in this study, we disable refraction at the interface while keeping the same shared-geometry two-media architecture.

3.2.2. Single-Sampler with Kinked Density

Rather than maintaining separate proposal-sampling hierarchies for air and water, BathyFacto uses a single proposal-network sampler that operates on a *virtual straight ray* spanning both media. For each camera ray that intersects the water surface at parameter t_I, we construct a virtual ray from the camera near plane to $t_I + t_{\text{far,water}}$, where $t_{\text{far,water}}$ is the maximum water-segment length obtained from the scene box intersection along the refracted direction. The proposal network samples along this virtual straight ray as if it were a single-medium scene.

To ensure that density is evaluated at physically correct positions despite the straight virtual ray, we wrap each proposal density function in a *kinked density* correction. For any sample at parameter t along the virtual ray, the wrapper computes the true 3D position as

$$\mathbf{x}(t) = \begin{cases} \mathbf{o} + t\,\mathbf{d} & \text{if } t \le t_I, \\ \mathbf{p}_I + \mathbf{d}_w\,(t - t_I) & \text{if } t > t_I, \end{cases} \tag{1}$$

where $\mathbf{p}_I = \mathbf{o} + t_I\,\mathbf{d}$ is the interface entry point and $\mathbf{d}_w$ is the refracted direction. The density network receives these corrected positions, so that the proposal distribution reflects the true scene geometry even though the sampler operates on a straight parameterization. This piecewise mapping, defined by Eq. (1), is also used for the final sample correction after proposal sampling converges: air samples remain on the original ray, while water samples are placed along the refracted direction.

In the experiments conducted in this study, we use two iterations of the proposal network with $(256, 96)$ proposal samples and 48 NeRF samples per ray (Table 1). Volume-rendering weights are computed from the original t-values on the virtual straight ray, preserving a consistent transmittance chain from camera to scene bottom.

3.2.3. Camera Pose Optimization

BathyFacto supports learnable camera pose refinement via an $\mathrm{SO}(3) \times \mathbb{R}^3$ camera optimizer. During training, the optimizer applies small rotation and translation corrections to each camera's ray bundle, reducing residual pose errors from the upstream SfM calibration. This is especially beneficial for bathymetric scenes where refraction-induced systematic errors may be coupled with pose uncertainty. The optimizer is regularized with L2 penalties on rotation ($\lambda_{\text{rot}} = 0.001$) and translation ($\lambda_{\text{trans}} = 0.01$).

3.2.4. Mask-Gated Two-Media Rendering

Water masks decide whether a pixel is treated as observing water or land. Only rays that both intersect the estimated water surface plane and are flagged by the medium mask extend the virtual ray through the water segment; non-water pixels remain air-only. In the no-refraction ablation, the virtual ray still extends through the interface, but the kinked density wrapper uses the original air direction instead of the refracted direction (refraction disabled). A configurable threshold (default: 0.5) binarizes the medium mask. Additionally, an optional per-pixel validity mask allows excluding "no data"-pixels or ignored regions from both training losses and evaluation metrics.

3.3. Two-media NeRF Interface Coupling

For each camera ray $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$, we compute the intersection parameter t_I with the planar water surface. Rays that do not intersect the plane within their near/far range are rendered as single-medium (air) rays. If a water mask indicates that a pixel falls within the water domain and the ray intersects the plane, we compute the interface entry point $\mathbf{p}_I = \mathbf{o} + t_I\mathbf{d}$ and the refracted direction below the surface. At the interface, we compute the refracted direction $\mathbf{d}_w$ using Snell's law with known refractive indices n_{air} and n_{water}. Let $\mathbf{n}$ be the unit normal of the planar interface and $\mathbf{d}$ the unit incident direction in air. We first orient $\mathbf{n}$ so that it opposes the incident ray by enforcing

$$\cos\theta_i = -\mathbf{n}^\top\mathbf{d} \ge 0, \tag{2}$$

flipping $\mathbf{n}$ when necessary. The direction $\mathbf{d}_w$ of the refracted underwater ray is then calculated by:

$$\mathbf{d}_w = \mathrm{refract}(\mathbf{d}, \mathbf{n}, n_{\text{air}}, n_{\text{water}})\,. \tag{3}$$

The resulting two-segment ray geometry is illustrated in Figure 4.

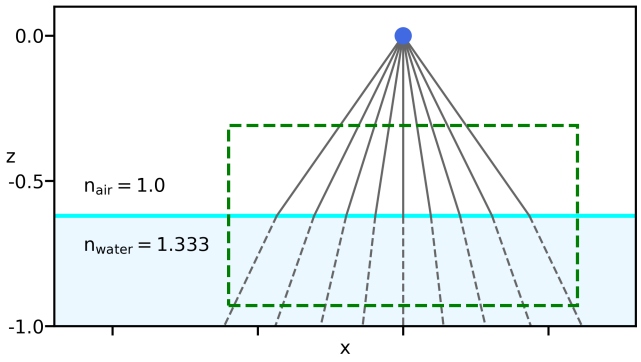

Figure 4: Two-segment ray geometry in the BathyFacto coordinate frame. Rays are shown in solid gray, the refracted segments below the water surface as dashed lines. The planar air-water interface separates the two media with their respective refractive indices $n_{air} = 1.0$ and $n_{water} = 1.333$. The green dashed rectangle denotes the scene box within which the single proposal sampler allocates samples via the kinked density wrapper.

We intersect the refracted ray with the scene box to obtain the maximum water-segment length $t_{far,water}$. A virtual straight ray is then constructed from the camera near plane to $t_I + t_{far,water}$, providing the single proposal-network sampler with a continuous parameterization across both media.

The kinked density wrapper (Section 3.2.2) ensures that density is evaluated at physically correct positions: air samples at $t \leq t_I$ are queried at $\mathbf{o}+t\mathbf{d}$, while water samples at $t > t_I$ are queried at $\mathbf{p}_I + \mathbf{d}_w(t-t_I)$. This correction is transparent to the proposal sampler, which perceives a standard single-medium ray.

After proposal sampling, the final NeRF sample positions and viewing directions are corrected analogously. Each sample is assigned a medium flag m_i: $m_i = 0$ for $t_i \leq t_I$ (air) and $m_i = 1$ for $t_i > t_I$ (water). The medium flag is passed to the color head for medium-conditioned appearance prediction.

Volume rendering uses the standard alpha-compositing formulation on the t-parameterized virtual ray. Given densities σ_i and sample intervals Δ_i, we compute

$$\alpha_i = 1 - \exp(-\sigma_i \Delta_i), \tag{4}$$

$$T_1 = 1, \qquad T_i = T_{i-1}(1-\alpha_{i-1}) \text{ for } i > 1, \tag{5}$$

$$w_i = T_i \, \alpha_i. \tag{6}$$

Because the virtual ray provides a monotonically increasing t-parameterization across both media, the transmittance chain naturally couples the air and water segments without requiring explicit sample concatenation or sorting. Consequently, underwater samples inherit the residual transmittance of the air segment rather than restarting with unit transmittance at the interface.

3.4. Deviations from Nerfacto

BathyFacto is built on the Nerfacto backbone within Nerfstudio (Tancik et al., 2023) and retains its hash-grid encoding, proposal-network sampling, and training infrastructure. It differs in three key aspects. First, BathyFacto introduces a physically constrained two-media ray model at a planar air-water interface with explicit Snell's-law refraction (optionally disabled for ablation), whereas Nerfacto assumes a single homogeneous medium. Second, BathyFacto conditions the color head on a one-bit medium flag (air or water) while sharing the density representation, whereas Nerfacto uses a single unified field without medium awareness. Third, BathyFacto wraps the proposal density functions in a kinked density correction that redirects water-segment positions along the refracted ray, enabling a single proposal hierarchy to allocate samples across both media, whereas Nerfacto uses its single proposal hierarchy without position correction. Scene contraction is disabled in all experiments, as the bathymetric scenes are bounded.

3.5. Training Losses

We supervise BathyFacto with a photometric RGB loss between the rendered image and the ground-truth image. For a set of pixels $\mathcal{I}$ with RGB values $\mathbf{C}_i$ and predictions $\hat{\mathbf{C}}_i$, we use an MSE loss,

$$\mathcal{L}_{rgb} = \text{mean}_{i\in\mathcal{I}} \left\| \mathbf{C}_i - \hat{\mathbf{C}}_i \right\|_2^2. \tag{7}$$

If an explicit validity mask is available (e.g., from an alpha channel, "no data"-pixels, or ignore regions), we exclude invalid pixels and re-normalize by the number of valid pixels to keep the loss magnitude comparable across batches. In the compared BathyFacto runs, medium masks are available for ray gating; the Nerfacto baseline is trained as a single-medium model without this two-media gating. Camera pose optimization adds an additional regularization loss that penalizes excessive rotation and translation corrections.

To reduce "floater" artifacts and encourage compact density along rays, we use the distortion regularizer (Mip-NeRF 360 style) on all samples along the virtual ray. When proposal networks are enabled, we additionally apply an interlevel loss that supervises the proposal sampling distribution of the single proposal hierarchy.

Overall, the training objective sums the photometric and regularization terms,

$$\mathcal{L} = \mathcal{L}_{rgb} + \lambda_{dist}\mathcal{L}_{dist} + \lambda_{inter}\mathcal{L}_{inter}, \tag{8}$$

with fixed coefficients ($\lambda_{dist} = 0.002$ and $\lambda_{inter} = 1.0$).

3.6. Refraction-Corrected Point Cloud Export

Standard NeRF point cloud export assumes straight rays and therefore produces biased underwater geometry when refraction is present. Therefore, we extend the exporter to back-project points along the two-segment ray geometry: for rays intersecting the air-water interface and reaching into water, we compute the interface point and project the

remaining depth along the refracted direction, while non-water rays are projected along the original direction.

In Nerfstudio, the Bathy point cloud exporter samples rays for all dataset images, evaluates the model, and back-projects one 3D point per ray using the predicted export depth. Rays are retained only if their accumulated opacity exceeds a fixed threshold, which suppresses background and other low-confidence samples. When refraction correction is enabled, the exporter uses the learned interface parameters (plane normal and plane intercept) together with the refractive indices to compute the per-ray refraction and applies the two-segment backprojection described above.

Beyond Nerfstudio's standard dataparser transform, the dataset pipeline stores reversible normalization metadata that map points from the normalized training frame back to the original photogrammetric coordinate system. Let $\mathbf{x}_{\text{norm}}$ denote a point in the normalized training frame, $\mathbf{R}_{\text{norm}}$ the rotation that aligns the fitted water-plane normal with the positive z-axis, $\mathbf{c}$ the centroid of the rotated camera positions used for centering, and s_{norm} the isotropic normalization scale. Using column-vector notation, the inverse normalization yields the point in the chunk-local photogrammetric frame:

$$\mathbf{x}_{\text{chunk}} = \mathbf{R}_{\text{norm}}^{\top}(s_{\text{norm}}\,\mathbf{x}_{\text{norm}} + \mathbf{c}), \tag{9}$$

where $\mathbf{x}_{\text{chunk}}$ denotes chunk-local coordinates. Let s_{chunk}, $\mathbf{R}_{\text{chunk}}$, and $\mathbf{t}_{\text{chunk}}$ denote the Metashape chunk transform that maps chunk-local coordinates to the original global frame. The final back-transformation is

$$\mathbf{x}_{\text{global}} = \mathbf{R}_{\text{chunk}}(s_{\text{chunk}}\mathbf{x}_{\text{chunk}}) + \mathbf{t}_{\text{chunk}}. \tag{10}$$

This enables direct metric comparison with reference geometry in the original global frame and preserves geometric consistency when returning exported point clouds from the normalized training space to the source photogrammetric frame.

4. Experimental Setup

This section summarizes the data, training setup, and evaluation protocol used in our experiments. We introduce the simulation dataset providing ground truth, describe implementation details and hyperparameters, list the compared configurations, and outline the metrics used to assess image quality and geometric accuracy.

4.1. Simulation Dataset (Blender)

The dataset foundational for this study is available for download (Schulte et al., 2026), with the comprehensive simulation pipeline documented in detail in Schulte et al. (2025). The pipeline starts with the derivation of camera trajectories based on predefined mission parameters, such as image overlap, Ground Sampling Distance, survey area extent, and camera model. In this case, a Point-of-Interest (POI) mission pattern was implemented, generating 130 rendered images along an oblique trajectory. The poses are then imported into Blender (V4.5) (Blender Development Team, 2026) via its Python interface to simulate a photorealistic coastal scene that comprises a ship, terrestrial features, and a subaqueous volume. To simulate underwater conditions, a water block characterized by a refractive index of $n = 1.333$ is implemented, assuming an ideal camera model with zero lens distortion. A significant advantage of this simulation framework is the ability to rapidly iterate physical and mission-contingent variables while maintaining an environment free from uncontrolled systematic biases.

4.2. Implementation Details & Hyperparameters

We evaluated a Nerfacto baseline and BathyFacto in two ablation variants (refraction enabled vs. disabled). Table 1 summarizes the key parameters.

4.3. Experimental Configurations

We compare the following configurations to assess the impact of explicit refraction modeling:

Baseline (Nerfacto): The single-medium Nerfacto configuration used as a standard baseline, with the same hash-grid encoding and training parameters but no two-media modeling.

BathyFacto (Refraction): Our two-media BathyFacto configuration with planar Snell refraction enabled, shared density field, and medium-conditioned color head.

Ablation (BathyFacto No-Refraction): The same BathyFacto configuration with refraction disabled (rays continue straight through the interface), used to isolate the benefit of refractive geometry from the shared two-media architecture.

MVS baseline: A conventional Multi-View Stereo point cloud without refraction correction, included as a reference for the traditional photogrammetric pipeline.

All NeRF-based configurations use $\mathrm{SO3}\times\mathbb{R}^3$ camera pose optimization during training.

4.4. Evaluation Metrics

To evaluate the quality of the 3D reconstruction, the assessment procedure described in Schulte et al. (2025) was applied. This study also has shown that 2D image-space metrics are not suitable for assessing geometric reconstruction quality.

The generated point clouds and the reference mesh were cropped to the central area of $50 \times 50\,m^2$ and filtered with a plane-based statistical outlier removal filter (neighbors = 10, $\sigma = 2.0$) (CloudCompare, 2026). Since point clouds are provided in a common reference frame, distances were evaluated directly in this coordinate system. Using the reference mesh, the cloud-to-mesh (C2M) distances were calculated for all reconstructed point clouds. From these distances, the mean and standard deviation were derived to quantify reconstruction accuracy.

The method from Seitz et al. (2006) was used to estimate completeness, and 19.8 million points ($\mathrm{GSD} = 0.01\,\mathrm{m}$) were uniformly sampled across the surface area of the reference mesh $\mathcal{P}_{\text{ref}}$, resulting in a denser representation of the target geometry than the reconstructed point clouds. Subsequently, cloud-to-cloud (C2C) distances between $\mathcal{P}_{\text{ref}}$ and each reconstructed point cloud $\mathcal{P}_i$ were computed. Reference

Table 1
Key configuration parameters of the compared NeRF models.

Parameter	Nerfacto	BathyFacto (No Refraction)	BathyFacto (Refraction)
Max iterations	100 000	100 000	100 000
Train rays / batch	4096	4096	4096
Eval rays / batch	4096	4096	4096
NeRF samples / ray	48	48	48
Proposal samples / ray	(256, 96)	(256, 96)	(256, 96)
Distortion loss mult.	0.002	0.002	0.002
Scene contraction	disabled	disabled	disabled
Camera optimizer	$SO3\times\mathbb{R}^3$[a]	$SO3\times\mathbb{R}^3$[a]	$SO3\times\mathbb{R}^3$[a]
Proposal schedule (`warmup/update/anneal`)	5000 / 5 / on	5000 / 5 / on	5000 / 5 / on
Proposal anneal max iters	1000	1000	1000
Hash-grid levels / max res	16 / 2048	16 / 2048	16 / 2048
Hidden dim (density / color)	64 / 64	64 / 64	64 / 64
Optimizer	Adam (lr 10^{-2})	Adam (lr 10^{-2})	Adam (lr 10^{-2})
Scheduler	Exp. decay $\rightarrow 10^{-4}$	Exp. decay $\rightarrow 10^{-4}$	Exp. decay $\rightarrow 10^{-4}$
Refractive indices $(n_{\text{air}}, n_{\text{water}})$	–	(1.0, 1.333)	(1.0, 1.333)
Media model	single field	shared density, 1-bit conditioned head	shared density, 1-bit conditioned head
Refraction model	none (single-medium)	disabled (ablation)	Snell refraction (enabled)

[a] Learnable rotation (SO3) and translation ($\mathbb{R}^3$) per camera, regularized with L2 penalties ($\lambda_{\text{rot}} = 0.001$, $\lambda_{\text{trans}} = 0.01$).

points within a distance threshold of $0.3\,\text{m}$ to the nearest reconstructed point were classified as reconstructed ($\mathcal{P}_{\text{rec}}$). Completeness is defined as:

$$\text{Completeness} = \frac{|\mathcal{P}_{\text{rec}}|}{|\mathcal{P}_{\text{ref}}|} \tag{11}$$

5. Results

This section presents the qualitative and quantitative results of our evaluation on the simulation dataset.

Figure 5 shows rendered RGB images and depth maps for three representative evaluation views. All three configurations produce visually plausible RGB renderings. The depth maps, however, reveal clear differences: BathyFacto with refraction produces smooth, spatially consistent depth predictions across the underwater region. The no-refraction ablation shows a visible depth discontinuity at the water surface, where the straight-ray model fails to account for the change in optical path length. The Nerfacto baseline exhibits floater artifacts on the seabed, visible as scattered depth outliers in the depth maps, indicating incomplete density convergence in the water column.

5.1. Quantitative Results

Table 2 reports the 2D image-space evaluation for the Nerfacto baseline, BathyFacto with refraction, and the BathyFacto no-refraction ablation.

Evaluation protocol. All methods are evaluated on the same `eval` split (13 images). For each image, predictions are computed and valid-pixel masks are applied consistently across all models to remove "no data"-pixels. PSNR, SSIM, and LPIPS are computed on the masked images and averaged over all evaluation views.

Table 2
2D image-space evaluation on the simulation `eval` split (mean over 13 images). Metrics are computed over all valid pixels.

Method	PSNR ↑	SSIM ↑	LPIPS ↓
Nerfacto	23.61	**0.476**	0.220
BathyFacto (Refr. OFF)	**25.57**	0.451	**0.183**
BathyFacto (Refr. ON)	22.59	0.354	0.245

BathyFacto without refraction achieves the highest PSNR (25.57 dB) and best LPIPS (0.183), while BathyFacto with refraction scores lowest (22.59 dB PSNR). This is expected: the refraction-enabled model renders underwater pixels along physically bent ray paths, which shifts apparent positions relative to the ground-truth images rendered with straight projection. The no-refraction variant, by contrast, keeps rays straight and can therefore match the image-space pixel layout more closely, at the cost of geometrically incorrect depth. Nerfacto achieves the highest SSIM (0.476), likely because its single-medium field does not introduce the medium-conditioned color head.

Interpretation note. As discussed in Section 4, 2D image-space metrics do not directly imply superior 3D geometric accuracy; a method may achieve high PSNR by fitting pixel

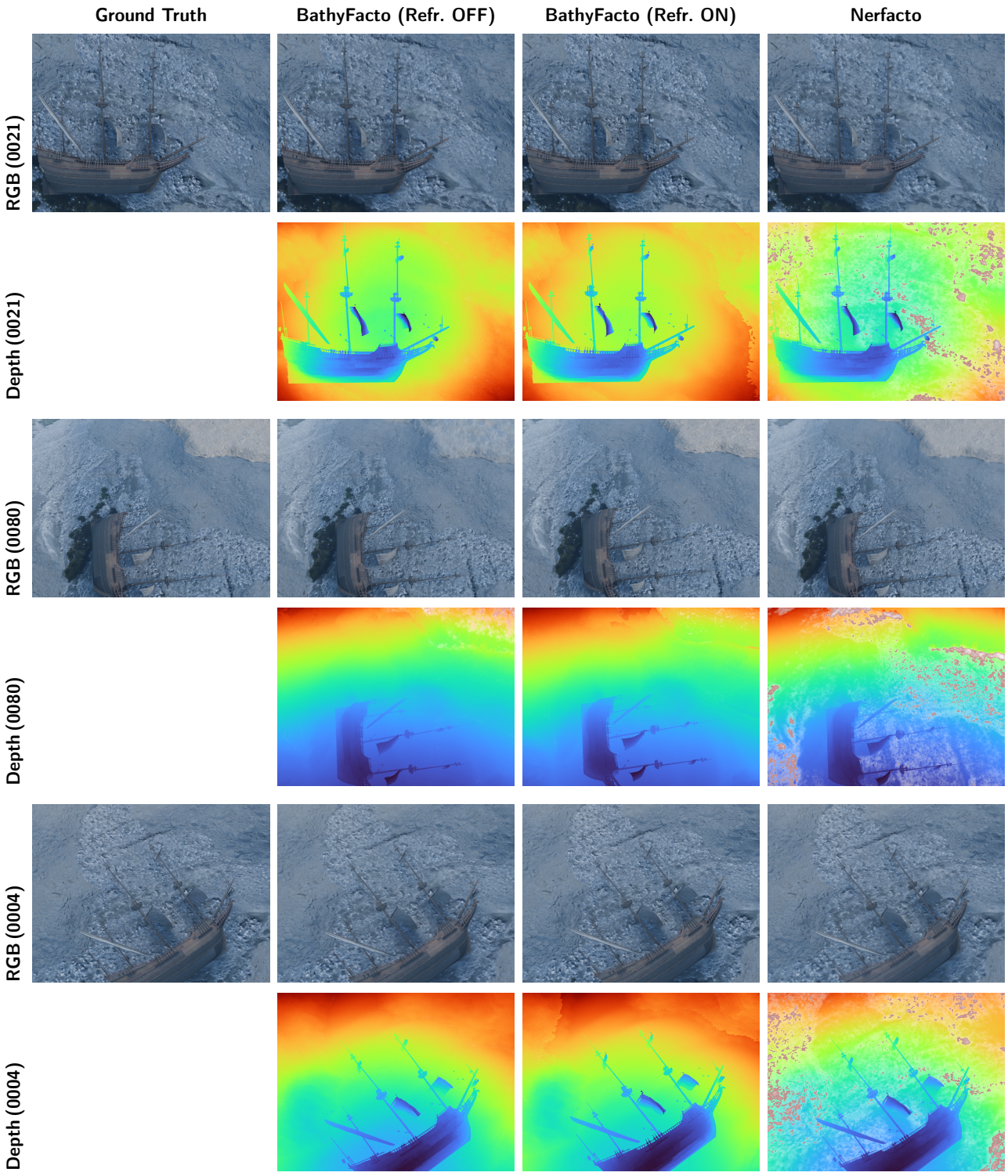

Figure 5: Qualitative 2D rendering comparison for three representative evaluation views. Columns: ground-truth RGB, BathyFacto without refraction (ablation), BathyFacto with refraction, and Nerfacto baseline. For each view, the top row shows RGB renderings and the bottom row shows predicted depth maps (no ground-truth depth available). Views span frontal (0021), steep-angle (0080), and oblique (0004) camera geometries.

colors with geometrically incorrect depth. The 3D point-cloud evaluation in the following section therefore remains the decisive criterion for geometric validation.

5.2. 3D Point-Cloud Evaluation

To directly evaluate geometric reconstruction quality, we compare the exported point clouds against the reference mesh in the common global frame. Cloud-to-Mesh (C2M)

Table 3
Quantitative 3D point-cloud evaluation in the common reference frame. Points with a C2M distance $> \pm 2$ m were filtered out before the calculation. Completeness is defined as the percentage of reconstructed points within ± 0.3 m of the reference mesh. In order to determine completeness meaningfully, the point clouds marked with * were first registered using ICP to the ship reference model (without the seabed). #Points [M] denotes the number of 3D points in millions.

Method	C2M Mean [m]	C2M Std [m]	Compl. [%]	#Points [M]
BathyFacto (Refr. OFF)*	0.43	0.54	30.0	3.9
BathyFacto (Refr. ON)	**0.06**	**0.17**	**87.0**	4.1
Nerfacto (no refraction)*	0.52	0.58	28.6	3.6
MVS (no refraction)*	0.36	0.80	20.6	1.7

signed distances, completeness, and a cross-section through the ship hull are reported in Figure 6, Figure 7, Figure 8, and Table 3.

Since the non-refracted point clouds (Nerfacto, BathyFacto Refr. OFF, MVS) exhibit a systematic depth offset due to the missing refraction correction, they were first aligned to the ship reference model (excluding the seabed) using a rigid-body ICP registration with six degrees of freedom (three rotations, three translations). BathyFacto with refraction enabled did not require ICP registration, as its point cloud already aligns well with the reference geometry. Points with a C2M distance exceeding ± 2 m were excluded before computing statistics.

5.3. Ablation Analysis

The 3D evaluation (Table 3) reveals distinct effects of refraction modeling:

Refraction modeling. BathyFacto with refraction enabled achieves the lowest C2M mean (0.06 m), the lowest standard deviation (0.17 m), and the highest completeness (87.0 %), indicating that explicit Snell's-law refraction at the interface produces geometrically more consistent and more complete underwater point clouds. The BathyFacto no-refraction ablation shows a substantially higher C2M mean (0.43 m) and lower completeness (30.0 %), confirming that the straight-ray model fails to reconstruct large portions of the underwater scene.

Single-medium baseline. The Nerfacto baseline (C2M mean 0.52 m, completeness 28.6 %) demonstrates that the single-medium model suffers from systematic depth bias without explicit refraction modeling.

Comparison to MVS. The traditional MVS pipeline without refraction correction (C2M mean 0.36 m) achieves reasonable accuracy on the reconstructed subset but with the lowest completeness (20.6 %), reflecting the limitations of straight-ray MVS in multimedia settings.

The cross-section through the ship hull (Figure 7) visually confirms that refraction-enabled BathyFacto traces the reference geometry more faithfully below the water surface.

6. Discussion

This section contextualizes BathyFacto in prior work and practical UAV bathymetry requirements. We summarize where explicit two-media modeling helps, discuss remaining limitations, and highlight implications for future datasets and modeling choices.

6.1. Contextualization and Comparison to Prior Work

BathyFacto builds on NeRFrac's foundational idea of embedding Snell's law into a NeRF pipeline (Zhan et al., 2023) but differs in several key aspects. First, BathyFacto uses a shared hash-grid-based density field with a medium-conditioned color head rather than fully separate air and water fields or a jointly optimized interface shape. This shared-geometry design enforces geometric consistency across the air-water boundary and converges substantially faster than sinusoidal-encoding approaches (100K vs. 200K+ iterations). Second, BathyFacto employs a single proposal-network sampler with a kinked density wrapper, which enables adaptive sample allocation across the air–water boundary without maintaining separate sampling hierarchies. Third, BathyFacto integrates directly into the Nerfstudio framework, inheriting its proposal-network sampling, appearance embeddings, and camera pose optimization—features not available in NeRFrac's vanilla-NeRF backbone. Fourth, BathyFacto provides a complete pipeline from photogrammetric input to metrically validated 3D point clouds in the original coordinate frame, which prior refractive NeRF research has not demonstrated.

Compared to classical through-water photogrammetry corrections (Agrafiotis et al., 2020; Lingua et al., 2023), BathyFacto operates on the volumetric representation level rather than correcting individual rays post-hoc, enabling joint optimization of geometry, appearance, and camera poses within a single framework.

6.2. Completeness evaluation

The completeness maps (Figure 8) visualize the spatial distribution of reconstructed points relative to the reference mesh. Colors encode the fraction of points within ± 0.3 m of the reference, while magenta areas labelled *nr* indicate reconstructed points which deviate by more than 0.3 m and are therefore counted as not reconstructed in the completeness metric. Methods marked with an asterisk (*) denote point clouds that were rigidly registered to the ship reference model using ICP prior to completeness evaluation. BathyFacto with refraction enabled achieves 87.0 % completeness, substantially outperforming both the no-refraction ablation (30.0 %) and the Nerfacto baseline (28.6 %). The MVS pipeline reaches only 20.6 %, reflecting the inherent limitations of straight-ray multi-view stereo in refractive environments.

6.3. Strengths and Limitations

Strengths. (i) The shared-geometry architecture with hash-grid encoding provides fast convergence and geometrically

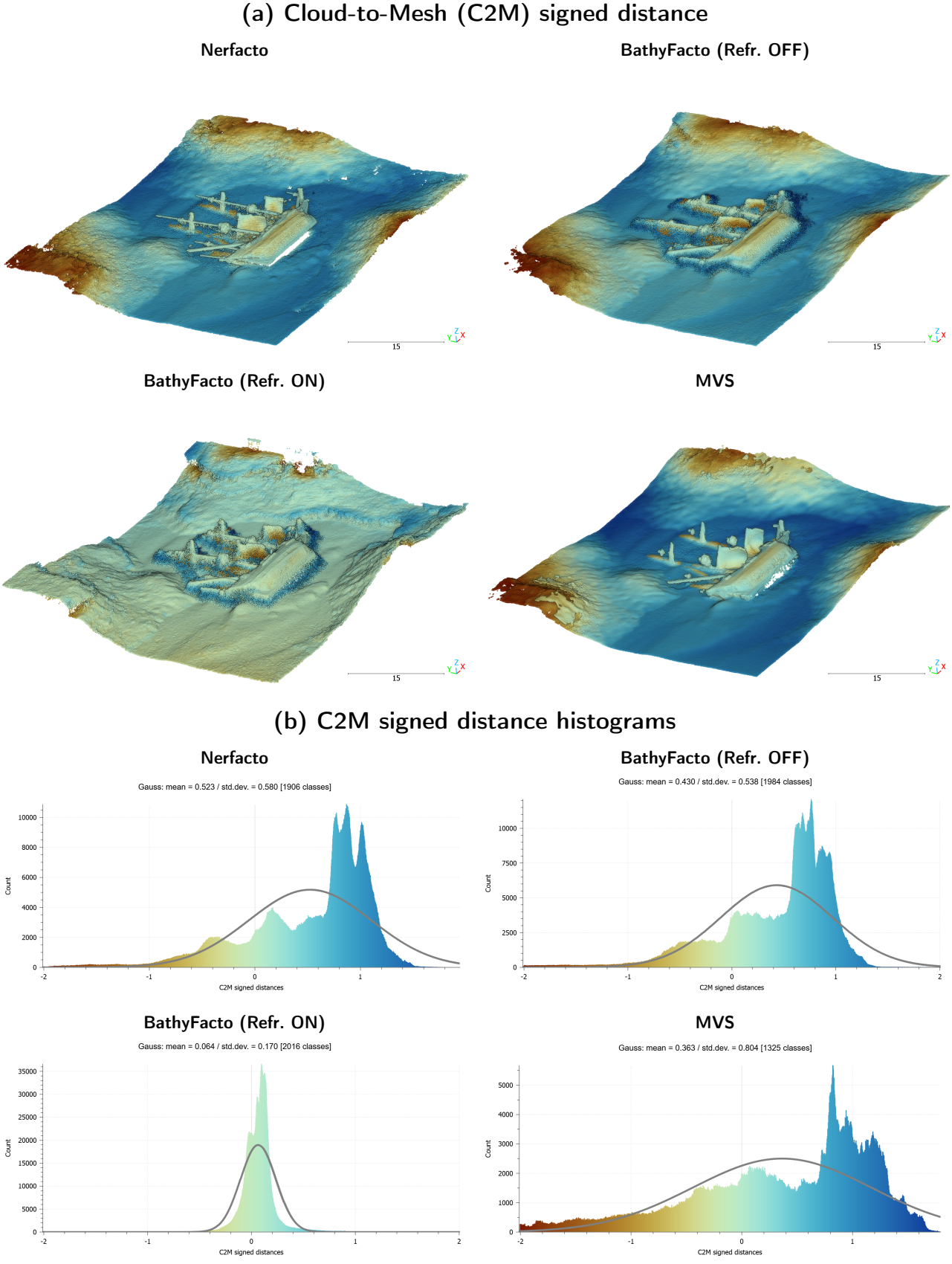

Figure 6: 3D point-cloud evaluation across all four configurations. **(a)** Cloud-to-Mesh (C2M) signed distance maps. BathyFacto with refraction enabled achieves the lowest C2M standard deviation (0.17 m). The color scale corresponds to the histograms shown in (b). **(b)** C2M signed distance histograms.

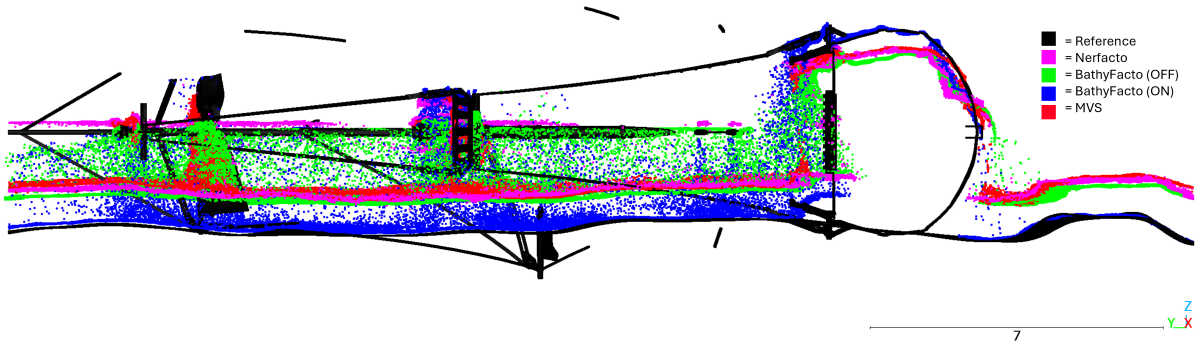

Figure 7: Cross-section through the ship hull. The refraction-enabled BathyFacto variant (blue) closely follows the reference geometry, whereas non-refracted variants exhibit a systematic depth error.

consistent two-media reconstruction. (ii) Nerfstudio integration ensures reproducibility and access to ongoing framework improvements. (iii) The refraction-corrected point cloud export uses the original photogrammetric camera poses and reversible coordinate transforms, enabling direct metric comparison with reference geometry in the original coordinate frame.

Limitations. (i) The water surface must be known a priori and is currently modeled as a plane; wavy or time-varying surfaces are not yet supported. (ii) Medium masks must be provided or generated in a preprocessing step; mask errors at shorelines can cause artifacts. (iii) The model has been validated only on synthetic data; real-world UAV imagery introduces additional challenges such as turbidity, caustics, and specular reflections.

7. Conclusion and Outlook

We have presented BathyFacto, a refraction-aware two-media extension of Nerfacto for photogrammetric bathymetry. BathyFacto uses a shared hash-grid-based density field with a medium-conditioned color head to represent air and water regions, traces camera rays as two piecewise linear segments connected by Snell's-law refraction at a planar water surface, and employs a single proposal-network sampler with a kinked density wrapper to allocate samples adaptively across both media. Integrated into Nerfstudio, the method provides a complete pipeline from photogrammetric SfM output to metrically validated 3D point clouds in the original coordinate frame.

On a synthetic two-media scene with known ground truth, BathyFacto with refraction enabled achieves 87.0 % completeness and a C2M standard deviation of 0.17 m, substantially outperforming both the single-medium Nerfacto baseline and the traditional MVS pipeline. The ablation with refraction disabled confirms that explicit two-media ray modeling is the primary driver of improved geometric completeness.

Future work will address the following directions: (i) extending the planar interface model to wavy or time-varying water surfaces, which is critical for real-world river and coastal applications; (ii) validating the method on real UAV imagery with ground-truth bathymetric references from airborne laser bathymetry; (iii) optimizing the loss function specifically for bathymetric accuracy rather than image-space fidelity; and (iv) exploring the reconstruction of submerged vegetation (macrophytes) as a challenging test case for volumetric underwater reconstruction.

Acknowledgements

The authors thank the Nerfstudio team for their open-source framework. Computations were performed on the GPU infrastructure of TU Wien.

Funding

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.

Declaration of Competing Interest

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

CRediT authorship contribution statement

Markus Brezovsky: Conceptualization, Methodology, Software, Validation, Writing - Original Draft. **Anatol Günthner:** Methodology, Visualization, Writing - Review & Editing. **Frederik Schulte:** Methodology, Validation, Writing - Review & Editing. **Lukas Winiwarter:** Methodology, Validation, Funding acquisition, Writing - Review & Editing. **Boris Jutzi:** Supervision, Resources, Funding acquisition, Writing - Review & Editing. **Gottfried Mandlburger:** Conceptualization, Supervision, Funding acquisition, Writing - Review & Editing.

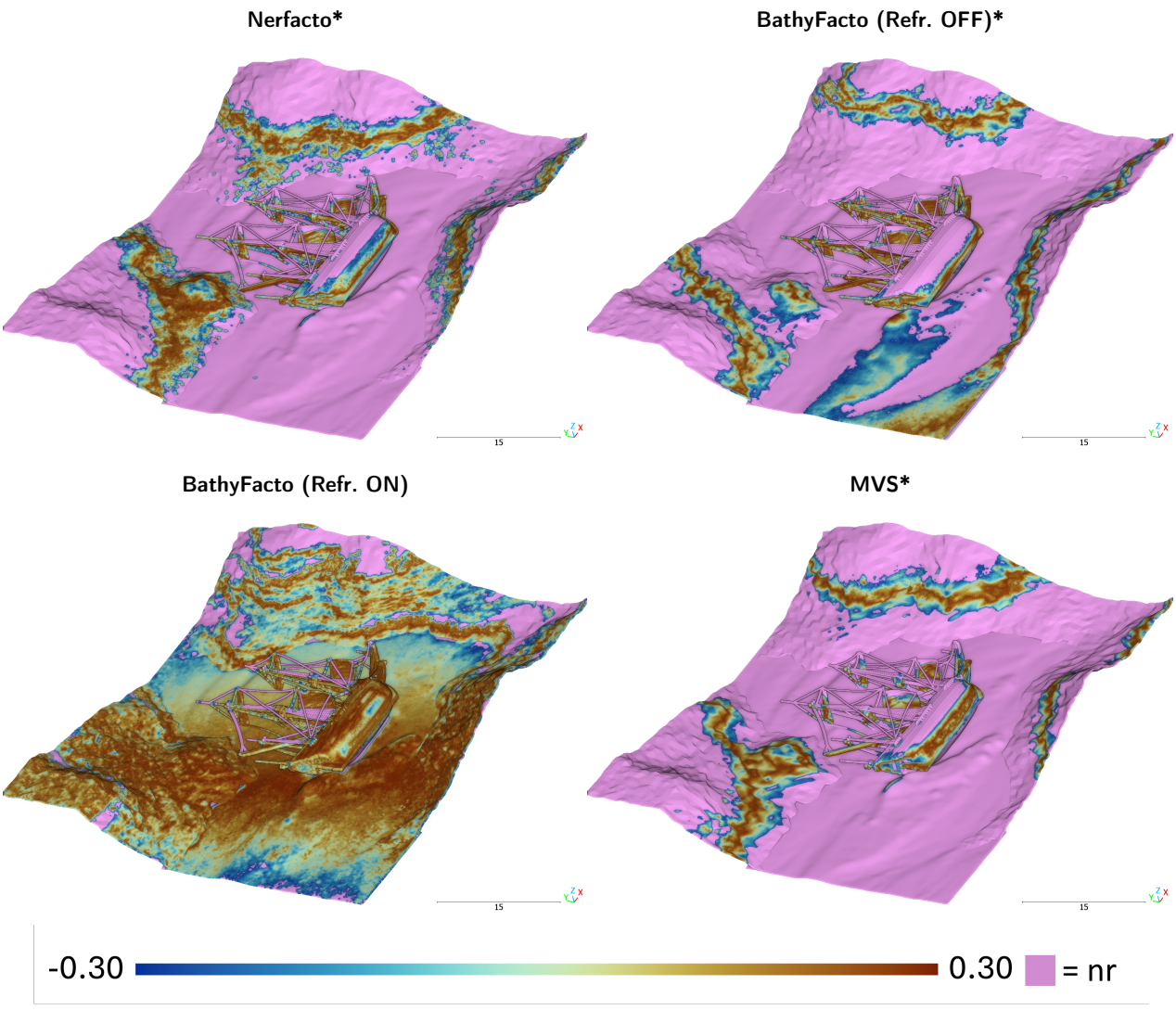

Figure 8: Completeness maps for all four configurations. Colors indicate the fraction of reconstructed points that lie within ± 0.3 m of the reference mesh in the common reference frame. Below the point cloud visualizations, a shared color legend is shown, mapping colors to local completeness values across all configurations. Magenta colored points, labelled *nr* for *not reconstructed*, do not lie within the ± 0.3 m tolerance to the reference mesh. Methods marked with an asterisk (*) were rigidly aligned to the ship reference model using ICP before completeness evaluation.

References

Agrafiotis, P., Karantzalos, K., Georgopoulos, A., Skarlatos, D., 2020. Correcting image refraction: Towards accurate aerial image-based bathymetry mapping in shallow waters. Remote Sensing 12, 322. doi:10.3390/rs12020322.

Barron, J.T., Mildenhall, B., Verbin, D., Srinivasan, P.P., Hedman, P., 2022. Mip-NeRF 360: Unbounded anti-aliased neural radiance fields, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5470–5479. doi:10.1109/CVPR52688.2022.00539.

Blender Development Team, 2026. Blender. version: 5.0 https://www.blender.org (16.02.2026). URL: https://www.blender.org.

Brezovsky, M., Günthner, A., Schulte, F., Winiwarter, L., Jutzi, B., Mandlburger, G., 2025. Analysis of Refraction Aware Neural Radiance Fields for 3D Reconstruction of through the Water Scenes. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLVIII-2-W10-2025, 39–45. doi:10.5194/isprs-archives-XLVIII-2-W10-2025-39-2025.

CloudCompare, 2026. CloudCompare (version 2.13.2). https://www.cloudcompare.org (16.02.2026). URL: https://www.cloudcompare.org/doc/wiki/index.php/Noise_filter.

Del Savio, A.A., Luna Torres, A., Vergara Olivera, M.A., Llimpe Rojas, S.R., Urday Ibarra, G.T., Neckel, A., 2023. Using UAVs and Photogrammetry in Bathymetric Surveys in Shallow Waters. Applied Sciences 13, 3420. doi:10.3390/app13063420.

Dietrich, J.T., 2017. Bathymetric structure-from-motion: extracting shallow stream bathymetry from multi-view stereo photogrammetry. Earth Surface Processes and Landforms 42, 355–364. doi:10.1002/esp.4060.

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. doi:10.1145/358669.358692.

Fujitomi, T., Hamaguchi, R., Onishi, M., Sakurada, K., 2022. LB-NeRF: Light bending neural radiance fields for transparent medium, in: Proceedings of the IEEE International Conference on Image Processing (ICIP), pp. 2142–2146. doi:10.1109/ICIP46576.2022.9897642.

Furukawa, Y., Hernández, C., 2015. Multi-view stereo: A tutorial. Foundations and Trends® in Computer Graphics and Vision 9, 1–148. doi:10.1561/0600000052.

Guenther, G., Cunningham, A., LaRocque, P., Reid, D., 2000. Meeting the accuracy challenge in airborne lidar bathymetry. Proceedings of EARSeL-SIG-Workshop LIDAR 1.

Günthner, A., Brezovsky, M., Schulte, F., Winiwarter, L., Mandlburger, G., Jutzi, B., 2025. Exploring the Potential of Refractive NeRFs 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 XLVIII-2-W10-2025, 107–114. doi:10.5194/isprs-archives-XLVIII-2-W10-2025-107-2025.

Hirschmuller, H., 2008. Stereo processing by semiglobal matching and mutual information. IEEE Transactions on Pattern Analysis and Machine Intelligence 30, 328–341. doi:10.1109/TPAMI.2007.1166.

Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G., 2023. 3D Gaussian Splatting for real-time radiance field rendering. ACM Transactions on Graphics 42, 139:1–139:14. doi:10.1145/3592433.

Kim, W., Fukiage, T., Oishi, T., 2023. REF^2-NeRF: Reflection and refraction aware neural radiance field. arXiv preprint arXiv:2311.17116 doi:10.48550/arXiv.2311.17116.

Levy, D., Peleg, A., Pearl, N., Rosenbaum, D., Akkaynak, D., Korman, S., Treibitz, T., 2023. SeaThru-NeRF: Neural radiance fields in scattering media, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 56–65. doi:10.48550/arXiv.2304.07743.

Lingua, A.M., Maschio, P., Spadaro, A., Vezza, P., Negro, G., 2023. Iterative refraction-correction method on MVS-SfM for shallow stream bathymetry, in: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, pp. 249–255. doi:10.5194/isprs-archives-XLVIII-1-W1-2023-249-2023.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60, 91–110. doi:10.1023/B:VISI.0000029664.99615.94.

Maas, H.G., 2015. On the accuracy potential in underwater/multimedia photogrammetry. Sensors 15, 18140–18152. doi:10.3390/s150818140.

Mandlburger, G., 2022. A review of active and passive optical methods in hydrography. The International Hydrographic Review 28, 8–52. doi:10.58440/ihr-28-a15.

Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R., 2020. NeRF: Representing scenes as neural radiance fields for view synthesis, in: Proceedings of the European Conference on Computer Vision (ECCV), pp. 405–421. doi:10.1007/978-3-030-58452-8_24.

Müller, T., Evans, A., Schied, C., Keller, A., 2022. Instant neural graphics primitives with a multiresolution hash encoding. ACM Transactions on Graphics 41, 102:1–102:15. doi:10.1145/3528223.3530127.

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) doi:10.1007/978-3-319-46487-9_31.

Schulte, F., Brezovsky, M., Günthner, A., Jutzi, B., Mandlburger, G., Winiwarter, L., 2025. Simulation and validation of underwater scenes for two-media optical 3d reconstruction XLVIII-2/W10-2025, 271–278. doi:10.5194/isprs-archives-xlviii-2-w10-2025-271-2025.

Schulte, F., Brezovsky, M., Günthner, A., Jutzi, B., Mandlburger, G., Winiwarter, L., 2026. Synthetic photogrammetric dataset for two-media 3d reconstruction: Shipwreck & terrain. doi:10.48323/G3CAA-ER166.

Schönberger, J.L., Frahm, J.M., 2016. Structure-from-motion revisited. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) , 4104–4113doi: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) , 519–528URL: https://doi.org/10.1109/CVPR.2006.19, doi: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, in: ACM SIGGRAPH 2023 Conference Proceedings, pp. 1–12. doi:10.1145/3588432.3591516.

Triggs, B., McLauchlan, P.F., Hartley, R.I., Fitzgibbon, A.W., 2000. Bundle adjustment — a modern synthesis, in: Goos, G., Hartmanis, J., Van Leeuwen, J. (Eds.), Vision Algorithms: Theory and Practice. Springer Berlin Heidelberg. volume 1883, pp. 298–372. doi:10.1007/3-540-44480-7_21.

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, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5481–5490. doi:10.1109/CVPR52688.2022.00541.

Westoby, M.J., Brasington, J., Glasser, N.F., Hambrey, M.J., Reynolds, J.M., 2012. 'Structure-from-Motion' photogrammetry: A low-cost, effective tool for geoscience applications. Geomorphology 179, 300–314. doi:10.1016/j.geomorph.2012.08.021.

Zhan, Y., Nobuhara, S., Nishino, K., Zheng, Y., 2023. NeRFrac: Neural Radiance Fields through Refractive Surface, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 18402–18412. doi:10.1109/ICCV51070.2023.01687.

Zhang, X., Zhang, Z., Ran, L., Li, X., 2026. Refractive neural rendering of underwater scenes via complex refractive surface reconstruction, in: Pattern Recognition and Computer Vision, pp. 262–276. doi:10.1007/978-981-95-5737-0_19.