

# A Comparative Analysis of Refraction-Aware SfM, Hierarchical Localization, and Gaussian Splatting for Underwater 3D Reconstruction

Fickrie Muhammad<sup>1,2</sup>, Rifakhryza A Mugaraya<sup>1,3</sup>, Gabriella Alodia<sup>2</sup>, Harald Sternberg<sup>1</sup>

<sup>1</sup> Hafencity University Hamburg, Hydrography and Geodesy Department, 20457, Hamburg, Germany –  
(fickrie.muhammad, rifakhryza.mugaraya, harald.sternberg)@hcu-hamburg.de

<sup>2</sup> Institut Teknologi Bandung, Hydrography Research Group, 40133, Bandung, Indonesia –  
(fickrie, gabriella.alodia)@itb.ac.id

<sup>3</sup> GEO Group, 26386, Wilhelmshaven, Germany – R.Mugaraya@geogroup.de

**Keywords:** Underwater Photogrammetry, 3D Reconstruction, Refraction Adjustment, Feature Matching, Hierarchical Localization (HLOC), Gaussian Splatting.

## Abstract

Underwater 3D reconstruction requires handling both geometric distortion and degraded visual conditions. This paper compares three complementary methods: a refraction-aware Structure-from-Motion (RSfM) pipeline using Underwater Colmap (UW-Colmap), a deep learning-based Hierarchical Localization framework (HLOC), and a neural rendering approach using Gaussian Splatting (GS). The first applies nonlinear refraction correction via a modified Colmap pipeline to compensate for distortions introduced by flat-pane housings. It improves geometric consistency and reduces artifacts in tank and open-water captures but relies on accurate refractive modeling. HLOC enhances matching robustness in low-contrast and low-texture scenes using SuperPoint and SuperGlue. However, it introduces considerable noise, particularly with retrieval and exhaustive matching, resulting in degraded reconstruction accuracy without geometric correction. Gaussian Splatting provides real-time rendering of visually realistic scenes using 3D Gaussian primitives. While not designed for structural accuracy, it delivers high visual quality when supplied with calibrated poses. The paper's core contribution is a controlled, side-by-side evaluation of these methods using a dual-environment dataset (air and underwater). By applying consistent evaluation metrics, geometry alignment, surface completeness, and visual consistency, we reveal the strengths and limitations of each approach. Results show that RSfM combined with GS provides the most reliable reconstruction and visualization pipeline. Deep learning methods are best applied at the feature level, followed by structured SfM for accurate geometry. This offers practical guidance for underwater photogrammetry and highlights the potential of hybrid reconstruction strategies.

## 1. Introduction

Reconstructing a three-dimensional (3D) model from underwater imagery presents a complex set of challenges not present in terrestrial photogrammetry. The underwater environment introduces multiple sources of distortion that affect image quality and geometric accuracy (Glaeser and Schröcker, 2000). Among the most significant issues are refraction at the camera housing interface, scattering and absorption of light, and lack of similar visual features due to poor textures and lighting conditions (Menna et al., 2017; Mandlbürger, 2022). These factors degrade the photometric and geometric signals critical for multi-view stereo and structure-from-motion (SfM) methods.

Refraction is the most fundamental problem in underwater photogrammetry. When light passes through a flat lens housing such as glass, it bends at the water-glass and glass-air interfaces, introducing nonlinear distortions (Jordt-Sedlazeck and Koch, 2012b). Traditional SfM pipelines assume an axial camera model, which is violated in this setting. As a result, pose estimation and 3D point triangulation can suffer from systematic errors. Several researchers have proposed solutions to this problem, including calibration techniques, multi-layer media models, or physics-based ray tracing through refractive layers (Gu et al., 2022; Jordt-Sedlazeck and Koch, 2012b; Łuczyński et al., 2017).

In parallel, the problem of low-quality visual features remains a challenge. Underwater scenes often lack high-contrast textures or

exhibit repetitive patterns like sand ripples, coral structures, or artificial artifacts. While robust to specific changes in viewpoint and lighting, traditional feature detectors such as SIFT and SURF tend to fail in scenes with heavy visual degradation (Ruble et al., 2014). This can lead to incorrect matches, reduced image overlap, and fragmented reconstructions. Various enhancement techniques have been proposed to mitigate this, including pre-processing steps like color correction, contrast enhancement, or dehazing (Zhou et al., 2022). Still, these methods typically operate on a single image and do not address geometric consistency across views.

Refractive Structure-from-Motion (RSfM) in Underwater Colmap (UW-Colmap) addresses the refraction distortion underwater by explicitly modelling how light bends through different media such as water-glass-air, resulting in more accurate camera poses and 3D reconstructions (She et al., 2024). This toolbox works by incorporating Snell's law of ray propagation into the optimization process. Meanwhile, deep learning-based methods such as Hierarchical Localization (HLOC) leverage learned keypoints and matchers, namely SuperPoint for feature detection and SuperGlue for matching, to establish correspondences that are more robust underwater (Sarlin et al., 2019).

Additionally, neural rendering techniques like Gaussian Splatting (GS) have emerged as an alternative way to visualize 3D content (Kerbl et al., 2023). These methods do not explicitly reconstruct dense geometry but render scenes directly from sparse point sets

and camera poses, encoding visual appearance in anisotropic Gaussian primitives (Li et al., 2025). Although not suited for metric reconstruction or measurement tasks, Gaussian Splatting can be valuable for producing high-quality visualizations from datasets where traditional SfM pipelines fail to generate complete models.

This paper presents a comparative study of three different approaches to underwater 3D reconstruction: refraction-aware structure-from-motion, deep feature-based localization and matching, and neural rendering through Gaussian Splatting. To ensure a fair evaluation, we introduce a dual-environment dataset that includes controlled captures of the same objects in air and underwater. By applying each pipeline to the same input data, we assess their performance in terms of geometric accuracy, reconstruction completeness, visual quality, and computational efficiency. Our goal is not only to highlight the strengths and limitations of each method but also to understand how they might be combined into hybrid systems that can better address the unique challenges of underwater photogrammetry..

## 2. Experimental Setup

All data were captured using a GoPro Hero 10, chosen for its compact size, wide-angle lens, and popularity in field-based underwater imaging. The camera was housed in a waterproof casing with 5 mm flat glass, introducing the refractive effects commonly encountered in practical deployments. The wide lens mode is selected because it has a large field of view of 105°. For data processing, a processing unit with an AMD Ryzen 9 5900X CPU, 64 GB RAM, and an NVIDIA RTX 3090 GPU is selected, ensuring the handling of computationally demanding tasks.

### 2.1 Data Acquisition

The multimedia camera setting is present because it is placed in a waterproof casing with 5 mm flat glass, introducing the refractive effects commonly encountered in practical deployments.



Figure 1 Prepared datasets featuring three objects: a Rubik's cube, a shipwreck model, and an aluminum cube with calibration patterns, captured in-air (left column) and underwater (right column).

For data acquisition, the camera captures continuous frames in video format at 1080p/30fps with Hypersmooth boost stabilization set off, allowing for telemetry recording and later for

motion stabilization IMU-based correction. To maintain consistent and controlled image sampling from video data, frames were extracted based on a predefined target frame rate. Depending on the original frame rate of the video, frames were selected at regular intervals to match this target rate. This approach ensured uniform spacing between frames across all datasets. Each selected frame was saved with a timestamp and label to support clear traceability throughout the reconstruction process.

Following image preparation, object selection was critical to evaluating reconstruction performance across various geometrical and visual challenges. Three physical objects, as seen in Figure 1, were chosen to represent increasing complexity in shape, scale, and texture. These objects allow for structured analysis of how each reconstruction pipeline performs under varied photogrammetric conditions, from fine detail in compact scenes to baseline accuracy in controlled setups.

Two types of environments were used to evaluate reconstruction under different spatial and visual constraints, as summarized in Table 1. The Rubik's cube and the shipwreck model were placed in a small plastic basin measuring approximately 60 × 40 cm with a depth of 20 cm. This setup simulates highly confined underwater scenarios where the camera must operate at close range with limited maneuverability. These conditions introduce practical challenges for photogrammetry, such as restricted viewing angles and shallow depth of field. In contrast, the large reference cube was placed in an indoor water tank with dimensions of 1.5 × 12.0 m and a depth of approximately 1.5 m.

Object	Dimensions (cm)	Purpose
Rubik's Cube	5 × 5 × 5	Tests fine detail capture in constrained views
Shipwreck Model	23.5 × 4 × 10.5	Simulates complex, irregular surfaces
Aluminium Cube	50 × 50 × 50	Tests larger objects with motion blur recorded

Table 1. Summary of Objects Used for 3D Reconstruction Evaluation

### 2.2 Calibration Data Processing

Accurate camera calibration is critical for reliable 3D reconstruction in underwater environments, where refraction through flat glass housings introduces significant distortions. Traditional calibration methods, such as submerging planar checkerboards, often fail to account for the nonlinear light paths caused by refraction at the water-glass-air interfaces, leading to geometric inaccuracies and projection errors, particularly in wide-angle configurations. This study employs a refractive-aware, object-based calibration approach using the CalibMar toolbox to address these challenges (CalibMar, Development Team, 2025). CalibMar is based on the refractive calibration model of flat port glass proposed by Jordt-Sedlazeck and Koch (2012), which simulates the nonlinear propagation of light across a multimedia model in the underwater environment.

The toolbox is adapted from She et al. (2024) to introduce the Refractive Structure-from-Motion (RSfM) in Underwater Colmap (UW-Colmap), integrating ray tracing modeling into the standard Colmap pipeline (Schönberger and Frahm, 2016). It highlighted practical challenges in underwater 3D modeling where refractive distortion undermines traditional triangulation (Chadebecq et al., 2020; Muhammad et al., 2025; Pedersen et al., 2018; Shortis, 2015), advocating for physically accurate

calibration models. Collectively, these works demonstrate the importance of modeling refractive geometry for improving the accuracy and robustness of underwater photogrammetric systems.

The calibration begins with determining the camera’s intrinsic parameters in air using standard pinhole modeling from images of a known 3D object (an aluminum cube with printed markers), with the calculated parameters in Table 2. These in-air parameters were then used to input the Calibmar toolbox, which incorporates refractive geometry, housing thickness, lens-to-glass distance, and interface normals, to simulate the light path through multimedia, assuming that the water properties are unchanged along the survey. Rather than re-optimizing the intrinsics underwater, Calibmar uses this information to generate a virtual camera model that more accurately represents underwater point projection, simulating the axial projection. Furthermore, the direct multiview calibration method estimates distortion directly from underwater image sequences, including refraction. However, this approach does not account for the physics of light bending at the interface and tends to add these errors to the distortion parameters.

Parameter	In-Air	UW	Description
fx	1132.130	1132.130	Focal length (x)
fy	1133.860	1133.860	Focal length (y)
cx	962.040	962.040	Principal point (x)
cy	544.365	544.365	Principal point (y)
k1	-0.261	-0.260	Distortion
k2	0.094	0.094	Distortion
p1	0.001	0.001	Distortion
p2	0.001	0.001	Distortion
Nx	–	0.012	Housing normal (x)
Ny	–	0.001	Housing normal (y)
Nz	–	0.999	Housing normal (z)
int_dist	–	-0.003	Lens-to-glass (m)
int_thick	–	0.005	Glass thickness (m)
na	–	1.000	Ref. index of air
ng	–	1.490	Ref. index of glass
nw	–	1.334	Ref. index of water

Table 2. In-air and underwater camera calibration parameters

### 3. Methods

This study evaluates three reconstruction pipelines for underwater photogrammetry under controlled conditions. The first is a refraction-aware Structure-from-Motion (RSfM) method using calibrated intrinsics and physical modeling of underwater light paths via UW-Colmap. The second integrates a hybrid pipeline in SfM with deep learning through Hierarchical Localization (HLOC). The third applies Gaussian Splatting, which synthesizes photorealistic views from sparse reconstructions using radiance-based 3D Gaussians.

#### 3.1 Underwater Photogrammetric Model

The underwater environment presents significant challenges to photogrammetry due to the refractive properties of water, which distort the light rays passing between media. Traditional SfM methods assume a single medium, typically air, and therefore struggle to accurately model the complexities of underwater imaging. To address this, UW-Colmap incorporates a refractive camera model and refractive SfM based on multimedial point projection (Agrawal et al., 2012) and refractive SfM (Jordt-Sedlazeck and Koch, 2012a). A notable alternative to the refraction calibration is the Pinax camera model (Łuczynski et

al., 2017), which addresses underwater refraction using a correction map beneficial for rectification and real-time applications such as SLAM. This model is based on a combination of a virtual pinhole and axial camera projection, allowing refraction correction through a real-time lookup table. However, unlike refractive SfM, which calculates point projection directly, the Pinax model does not account for view-dependent light path variation, making it less suited for large-scale reconstructions.

The refractive camera models in UW-Colmap explicitly consider the bending of light as it transitions between water, the camera housing, and air. In addition, two camera housings typical for multimedia photogrammetry are being tested: flat ports and dome ports. For flat ports, light undergoes refraction as it passes through the flat glass of the camera housing, altering the perceived positions of scene points. In the case of dome ports, the curved surface of the housing presents a more complex refraction geometry, where the refractive axis remains a crucial component in modeling the distortion. Unlike traditional SfM, which assumes linear epipolar geometry, refractive SfM introduces a feature-dependent virtual epipolar geometry to account for light refraction. The relative point projection and epipolar line between camera frames are readjusted by introducing a virtual camera correlated with the refraction index and Snell’s laws of ray propagation, correcting invalid axial projection underwater (She et al., 2024).

#### 3.2 Hierarchical Localization (HLOC)

Hierarchical localization (HLOC) is initially used to estimate camera poses in large environments based on the reconstructed scene or known environment (Sarlin et al., 2019). It starts with global image retrieval and then moves to local feature matching. The first step, coarse localization, uses global descriptors such as NetVLAD to select a small number of candidate images from a set of image scenes. NetVLAD combines the Vector of Locally Aggregated Descriptors (VLAD) aggregation method with a convolutional neural network (CNN). This allows it to create compact global descriptors that handle changes in viewpoint and lighting conditions across all scenes. In the fine localization step, local features are extracted from the query image and matched using deep learning models like SuperPoint (DeTone et al., 2018). These models use convolutional networks to extract distinctive and reliable features, even when scenes have little texture or are visually degraded. The matched features are then used to estimate the camera pose using geometric techniques, often with the Perspective-n-Point (PnP) algorithm combined with RANSAC to filter out incorrect matches.

Feature matching can be improved using more advanced neural network-based matchers such as SuperGlue (Sarlin et al., 2020). It limits and enhances fine matching to only the candidate images selected earlier, making the method efficient and scalable to larger datasets. At the same time, it maintains a high level of accuracy. Its modular design also makes it easy to swap in different tools for feature extraction or matching, which helps when working in varied or challenging environments. The global bundle adjustment is then used for 3D reconstruction in the pipeline, which helps refine the overall pipeline model for the entire database. Although global bundle adjustment aims to refine the reconstruction, it can introduce inaccuracies when initial errors or mismatches exist. To achieve a more robust reconstruction, feature matches from HLOC are integrated into the Colmap pipeline (Schönberger and Frahm, 2016). By combining deep learning-based matching from HLOC with Colmap’s incremental pose estimation, this hybrid approach

ensures resilient correspondences and improves reconstruction accuracy, as shown in Figure 2.

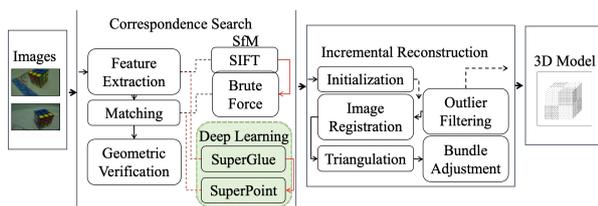


Figure 2. HLOC+Colmap pipeline.

### 3.3 Gaussian Splatting

Gaussian Splatting is used in the method to synthesize photorealistic views from sparse three-dimensional point clouds obtained through Structure from Motion (SfM) pipelines such as Colmap or HLOC (Sarlin et al., 2019). The process begins by estimating camera poses and sparse geometry from multiview images using Colmap or HLOC pipeline in correspondence searching. These outputs, including camera intrinsics, extrinsics, and 3D sparse points, serve as input to the GS pipeline. Each point in the reconstruction is represented as an anisotropic Gaussian, with attributes learned from the dataset. These attributes include camera pose, orientation, and color information. It directly rasterizes Gaussians in screen space. This approach enables faster rendering and is more robust to noise (Kerbl et al., 2023).

Initially, the parameters of each Gaussian are optimized using photometric consistency across input views. The known camera poses guide the projection of Gaussians, and their attributes are refined to minimize the difference between rendered and observed images, which is beneficial for underwater data. To address underwater environment challenges, some implementations introduce additional learning strategies. For example, WaterSplatting (Li et al., 2025) modifies the pipeline to enable fast reconstruction in underwater environments, incorporating techniques such as backscatter modeling and color correction. While it is not intended for precise geometric measurements, Gaussian Splatting is effective for qualitative visualization.

## 4. Results and Discussions

### 4.1 Feature Matching

In the UW-Colmap setup, SIFT Feature matching is used an exhaustive strategy to compare all image pairs. While thorough, this method often produces incorrect matches, particularly in scenes with repeated structures or low texture, such as the Cube dataset. HLOC takes a different approach using learned descriptors like SuperPoint and feature matchers like SuperGlue. Although HLOC can also perform exhaustive matching, its use of deep learning makes the results more selective. In practice, the number of false matches is lower than with traditional SIFT-based methods. This difference becomes clear when inspecting matched image pairs. In UW-Colmap, incorrect correspondences appear frequently, especially in complex or repetitive scenes. HLOC, by comparison, tends to produce more consistent and focused matches, shown in Figure 3.

The way images were captured also plays a role. Because the image sequences have a strong overlap between consecutive frames, most reliable matches are found along the diagonal of the

matching matrix, as seen in Figure 4. Matches outside this diagonal represent less overlapping image pairs, where errors are more likely.

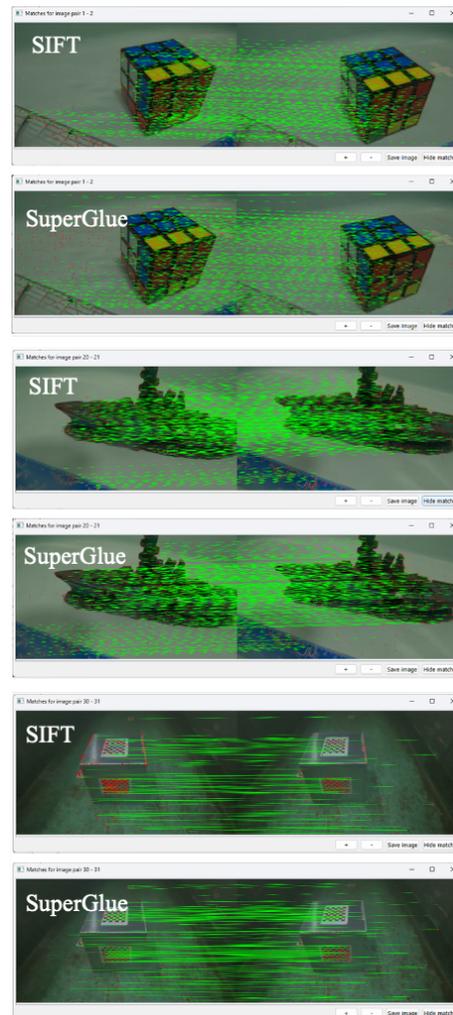


Figure 3. Feature Matching comparison between SIFT and SuperGlue in Rubik (top), Shipwreck (middle), and Cube (bottom) datasets.

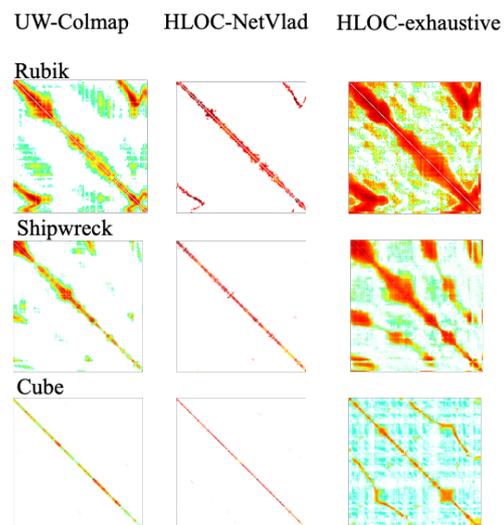


Figure 4. Match matrix comparison between UW-Colmap, HLOC-NetVlad, and HLOC-exhaustive for Rubik, Shipwreck, and Cube datasets.

Both methods show strong diagonal structures in the Rubik's dataset, suggesting that frame-to-frame matching works well. Colmap, however, also shows more matches outside that region, which often include mismatches due to similar-looking textures across views. HLOC keeps the matrix tighter around the diagonal, focusing only on image pairs with meaningful overlap. The Shipwreck dataset is more difficult due to uneven surfaces and weaker texture cues. Both methods show consistent diagonal patterns, but Colmap produces more off-diagonal matches that are less reliable. On the other hand, HLOC concentrates on feature matches in areas with actual image overlap, illustrating more consistent feature matching. Both tools perform well for the Cube dataset, which contains regular geometry and clear markers. Still, Colmap shows occasional off-diagonal connections, while HLOC's match structure remains cleaner.

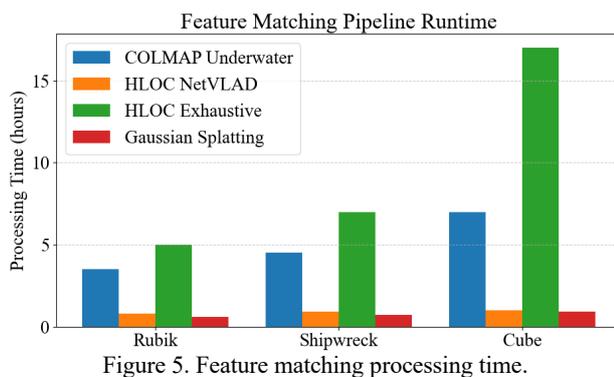


Figure 5. Feature matching processing time.

The way features are matched clearly affects processing time, as seen in Figure 5. The UW-Colmap here uses exhaustive matching, clearly showing long processing time, but the most consistent reconstruction. Using HLOC with exhaustive matching reduced the mismatched features, but took significantly longer. For example, the Rubik's Cube dataset required about 5 hours, the Shipwreck around 7 hours, and the Cube dataset took up to 17 hours. This reflects the high computational cost of matching many image pairs in detail. In comparison, HLOC and NetVLAD retrieval completed all datasets in less than an hour. This approach is much faster because it narrows down image pairs before matching. However, the results were sometimes less complete, and in some cases, incorrect matches affected the final reconstruction quality.

Gaussian Splatting processed all datasets in less than an hour because it avoids traditional feature matching and reconstruction steps. It uses precomputed camera poses and renders scenes directly from sparse point clouds, which makes the process significantly faster. While it's not intended for tasks like measurement or geometric evaluation, it does provide high-quality visualizations in a short amount of time. For comparison, the UW-Colmap setup generally completed faster than the HLOC Exhaustive pipeline. Depending on the dataset, total runtimes ranged from around 3 to 7 hours. Although HLOC can offer more robust feature matching under challenging conditions, Colmap balances speed and geometric accuracy, making it a practical option for structured underwater reconstruction tasks.

In the Rubik dataset, traditional SIFT-based matching frequently results in incorrect correspondences between non-overlapping views, indicated in Figure 6. The repetitive high-contrast color patterns of the Rubik's cube make it particularly prone to false matches, as similar-looking features, such as square corners and color transitions, are detected across multiple unrelated frames. These errors highlight SIFT's sensitivity to the lack of features in

all image sets. In contrast, SuperGlue demonstrates stronger feature matching, avoiding false feature candidates even in exhaustive matching.

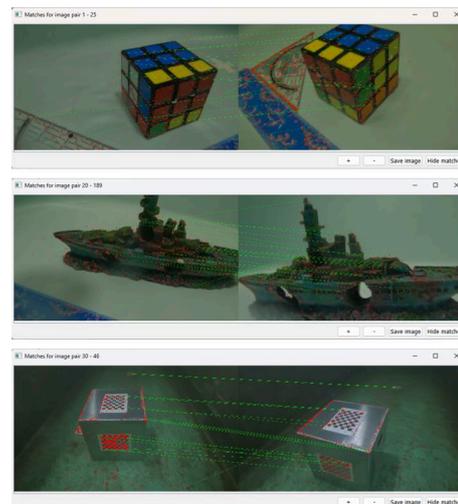


Figure 6. Typical mismatched features from UW-Colmap.

Matching features in the Shipwreck dataset is especially challenging because of the repeated structures and irregular geometry. Elements like railings or windows often appear very similar across different parts of the scene, which can confuse matchers that rely only on local descriptors. This sometimes leads to incorrect matches between frames that are far apart. SuperGlue tends to handle this better. The Cube dataset, while simpler in structure, also presents matching issues. Each face of the cube includes calibration markers with very similar textures, which makes it hard for local descriptors to tell one side from another. Overall, the results across datasets highlight a key difference between traditional and learned matching. SIFT is more susceptible to problems in scenes with repeated patterns or structured elements, while SuperGlue's ability to model global context makes it more robust.

## 4.2 3D reconstruction

The 3D reconstruction comparison evaluates results from UW-Colmap and HLOC combined using Colmap for dense reconstruction and neural rendering via GS, with Table 3 providing a summary of the results.

Object	Method	Result
Shipwreck Model	UW-Colmap	Sharp peak, minimal surface noise, highly consistent model
	HLOC (NetVLAD)	Noisy and broad error and noise distribution
	HLOC + Colmap	Moderate noise, no duplication artifacts, improved geometric
	GS	Sparse and inconsistent geometry, visually coherent
Rubik's Cube	UW-Colmap	Very narrow error distribution, clean reconstruction
	HLOC (NetVLAD)	Comparable to UW-Colmap , with noise increase
	GS	Visually appealing output, but sparse point density
Aluminium Cube	–	Failed due to motion blur and lack of detectable features

Table 3. Summary of reconstruction result

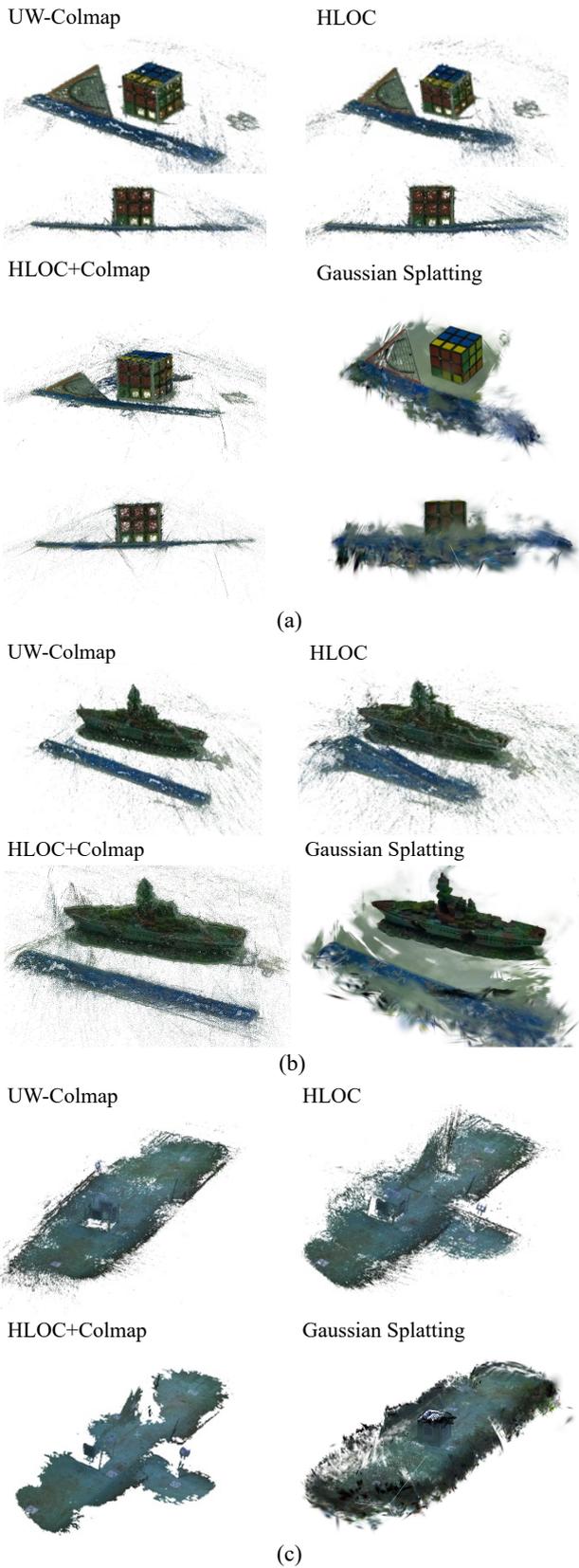


Figure 7. 3D reconstruction of (a) Rubik (b) Shipwreck, and (c) Cube dataset.

The dense reconstruction is then applied to all datasets, with the result shown in Figure 7. In the Rubik dataset, UW-Colmap yielded clean and geometrically accurate surfaces, with minimal noise around the cube and rulers. HLOC reconstructions

appeared slightly denser but introduced mild noise in regions of fine detail. Despite these issues, both pipelines effectively reconstructed the high-contrast, feature-rich Rubik model.

In the Shipwreck dataset, UW-Colmap maintained structural consistency across complex geometry. HLOC reconstructions showed more irregularities, especially in dense or ambiguous regions. Gaussian Splatting produced compelling visual results. It was particularly effective in rendering fine-scale features and providing a realistic appearance. However, at close range, its Gaussian splatter affected object edges and influenced the perception of the model's true shape. Along the model reconstruction, the Rubik and Shipwreck models lacked background geometry due to the low feature content of the surrounding regions. In contrast, the Cube dataset included pool walls and floors with enough texture for feature detection. Nevertheless, the reflective surface of the Cube model limited feature matching, causing reconstruction failure in all HLOC-based methods, with only UW-Colmap successfully producing a complete model.

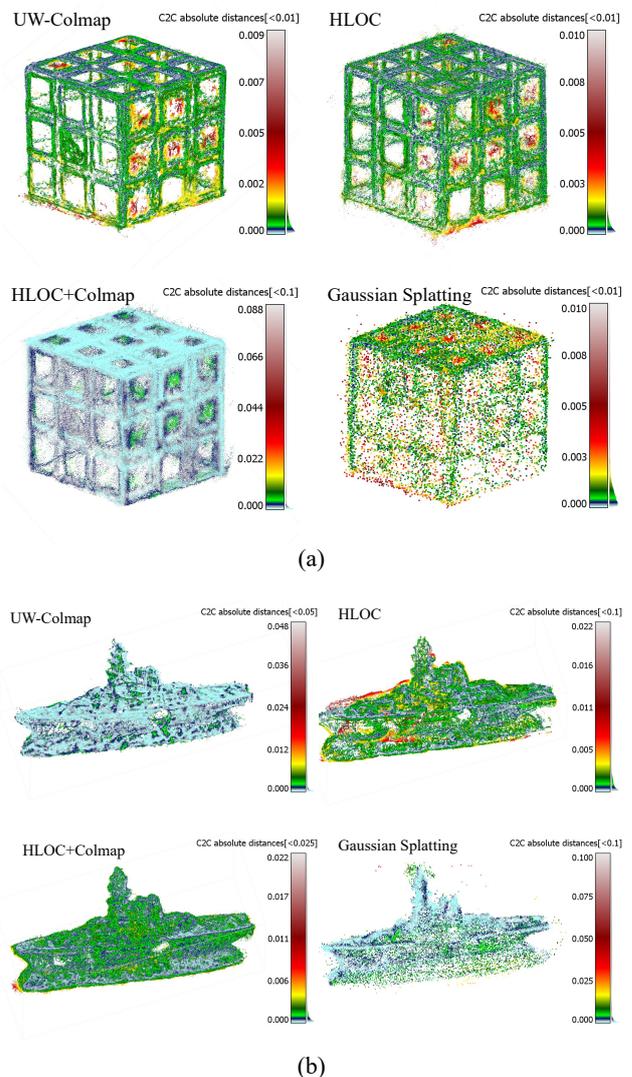


Figure 8. Cloud2cloud distance between in-air reconstruction and underwater reconstruction of (a) Rubik and (b) Shipwreck.

Due to its direct refraction modeling, UW-Colmap consistently delivered the most geometrically reliable outputs, while GS

excelled in photorealistic modelling. HLOC, combined with exhaustive matching and Colmap reconstruction, provided competitive results but remained more noisy. The accuracy of each reconstruction pipeline was assessed by comparing the underwater models to their corresponding in-air reconstructions using Cloud-to-Cloud (C2C) absolute distance metrics, shown in Figure 8. These distributions reflect how closely each method approximates the ground truth geometry and help quantify the presence of surface noise and misalignment. For the Shipwreck dataset, UW-Colmap produced the most accurate and consistent reconstruction. It yielded approximately 2 million points and a root mean square error (RMSE) of 1.8 mm, with a tightly concentrated error distribution and minimal surface deviation. This confirms that direct refractive modeling enhances geometric consistency in underwater environments.

HLOC using NetVLAD retrieval exhibited significantly more noise, including structural duplication—notably, a duplicated ship bow—resulting in a degraded model with an RMSE of 2.5 mm and fewer than 900,000 usable points. This indicates poor feature matching and the absence of native refraction correction, which are critical weaknesses in this configuration. However, when HLOC was used with exhaustive matching and its features were passed to Colmap for reconstruction, the results improved considerably after applying a correction map based on the Pinax model. This hybrid pipeline generated approximately 2.08 million points with an RMSE of 2.1 mm, effectively avoiding structural artifacts and reducing surface noise. However, it still falls short of the precision achieved by direct ray-traced modeling.

All methods performed well in the Rubik dataset, which features high contrast and strong geometry. UW-Colmap achieved around 2 million points with an RMSE of 1.5 mm, offering the highest precision. The HLOC+Pinax+Colmap combination produced approximately 920,000 points with an RMSE of 1.9 mm, showing good accuracy despite lower point density. Gaussian Splatting generated a visually coherent but sparse model (~17,000 points, RMSE 4.3 mm), demonstrating usefulness for visualization but not for metric analysis. The Cube dataset posed the most significant challenge. Only UW-Colmap succeeded in reconstructing a usable 3D model. HLOC-based pipelines failed due to motion blur, specular surfaces, and a lack of consistent visual features. These factors disrupted feature matching and triangulation, rendering the outputs incomplete or invalid for comparison. No C2C distance evaluation could be conducted for these failed cases, highlighting the robustness of direct refractive correction in degraded visual conditions.

In summary, UW-Colmap consistently produced the most accurate and dense reconstructions across all datasets. The HLOC+Colmap hybrid approach proved a viable alternative when equipped with exhaustive matching and correction via the Pinax model, though it remained sensitive to feature quality. Native HLOC with NetVLAD struggled in underwater environments, especially in scenes with limited visibility or ambiguous features. Gaussian Splatting was effective for visualization but not appropriate for high-precision reconstruction. These findings highlight the value of directly integrating refractive modeling into the reconstruction pipeline and the importance of reliable feature matching in challenging underwater photogrammetry.

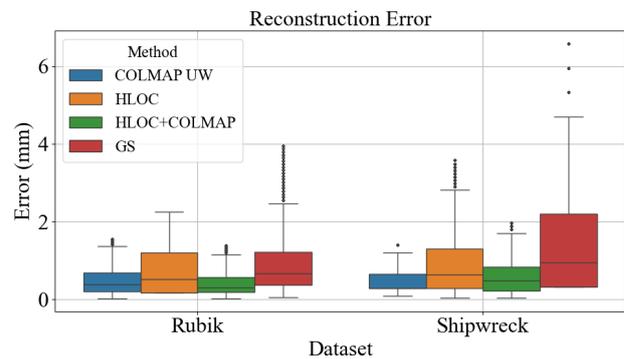


Figure 9. Reconstruction Error of Rubik and Shipwreck datasets.

Figure 9 compares the reconstruction error, in millimeters, for four different methods, UW-Colmap, HLOC, HLOC+Colmap, and GS, applied to two datasets: Rubik and Shipwreck. Each dataset is shown on the horizontal axis, and the vertical axis represents the error magnitude. The colored boxes indicate the range of error values for each method, with the thick horizontal line inside each box representing the median error. Thin lines (whiskers) extend to show the general spread of the data, while isolated dots indicate individual outlier values. In the Rubik's dataset, all methods show relatively low reconstruction errors. UW-Colmap and HLOC+Colmap have the smallest spread and lowest medians, suggesting more accurate and consistent results. HLOC shows slightly more variability, while GS has a visibly wider distribution, indicating less precision, though its overall errors remain low.

In the Shipwreck dataset, the differences between methods become more pronounced. UW-Colmap again shows the lowest error range, indicating it handles underwater reconstruction well. HLOC has a broader spread and more outliers, suggesting inconsistent performance. HLOC+Colmap improves on native HLOC, reducing error spread and lowering the median. GS shows the most extensive spread in this dataset, with high variability and multiple outliers, reflecting challenges in maintaining accurate geometry despite producing photorealistic outputs.

## 5. Conclusion

This study compares traditional and modern 3D reconstruction methods for underwater photogrammetry using a structured workflow built with open-source tools. It evaluates refraction-aware Structure-from-Motion (RSfM) using UW-Colmap, Hierarchical Localization (HLOC), and Gaussian Splatting using data from a GoPro Hero 10. Refractive calibration was performed using a physically based model and a submerged checkerboard, which were applied without further optimization to isolate their geometric effects. Because HLOC does not directly incorporate refractive modeling, calibration was applied through a correction map generated by the Pinax model to compensate for refractive distortions before reconstruction.

Experiments showed that refraction-aware SfM consistently produced accurate and clean reconstructions. HLOC improved feature matching but introduced noise, particularly in low-visibility, small-scale scenes. Combining HLOC's deep feature extraction with Colmap's reconstruction pipeline improved stability and achieved millimeter-level accuracy in the Rubik and Shipwreck datasets. However, in the Cube dataset, this approach struggled due to reflective surfaces and low texture, resulting in incomplete reconstruction and residual noise. Gaussian Splatting

provided visually compelling renderings and enhanced scene interpretability, especially in noisy or texture-poor underwater conditions. The future direction of the work will explore integrating deep learning directly into refractive modeling through learned ray tracing or data-driven depth correction to improve reconstruction robustness and accuracy in complex or large-scale underwater environments.

## References

- Agrawal, A., Ramalingam, S., Taguchi, Y., Chari, V., 2012: A theory of multi-layer flat refractive geometry Supplementary Materials. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 3346–3353.
- CalibMar, Development Team, 2024: CalibMar—Underwater Camera Calibration Toolbox. Christian-Albrechts-Universität zu Kiel. Available at: <https://cau-git.rz.uni-kiel.de/inf-ag-koeser/calibmar> (20 January 2025).
- Chadebecq, F., Vasconcelos, F., Lacher, R., Maneas, E., Desjardins, A., Ourselin, S., Vercauteren, T., Stoyanov, D., 2020: Refractive Two-View Reconstruction for Underwater 3D Vision. Int. J. Comput. Vis., 128, 1101–1117. <https://doi.org/10.1007/s11263-019-01218-9>.
- DeTone, D., Malisiewicz, T., Rabinovich, A., 2018: SuperPoint: Self-supervised interest point detection and description. In: CVPR Workshops, 224–236.
- Glaeser, G., Schröcker, H.P., 2000: Reflections on refractions. J. Geom. Graph., 4, 1–18. <https://doi.org/10.2307/3975136>.
- Gu, C., Cong, Y., Sun, G., Gao, Y., Tang, X., Zhang, T., Fan, B., 2022: MedUCC: Medium-Driven Underwater Camera Calibration for Refractive 3-D Reconstruction. IEEE Trans. Syst. Man, Cybern. Syst., 52, 5937–5948. <https://doi.org/10.1109/TSMC.2021.3132146>.
- Jordt-Sedlazeck, A., Koch, R., 2012a: Refractive structure-from-motion on underwater images. IEEE Int. Conf. Comput. Vis., 57–64. <https://doi.org/10.1109/ICCV.2013.14>.
- Jordt-Sedlazeck, A., Koch, R., 2012b: Refractive calibration of underwater cameras. Lect. Notes Comput. Sci., 7576, 846–859. [https://doi.org/10.1007/978-3-642-33715-4\\_61](https://doi.org/10.1007/978-3-642-33715-4_61).
- Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G., 2023: 3D Gaussian Splatting for Real-Time Radiance Field Rendering. ACM Trans. Graph., 42(6), 1–14. <https://doi.org/10.1145/3592433>.
- Lee, Z., Hu, C., Shang, S., Du, K., Lewis, M., Arnone, R., Brewin, R., 2013: Penetration of UV-visible solar radiation in the global oceans: Insights from ocean color remote sensing. J. Geophys. Res. Ocean., 118, 4241–4255. <https://doi.org/10.1002/jgrc.20308>.
- Li, H., Song, W., Xu, T., Elsig, A., Kulhanek, J., 2025: WaterSplatting: Fast Underwater 3D Scene Reconstruction Using Gaussian Splatting. In: Int. Conf. 3D Vision.
- Łuczyński, T., Pflingstorn, M., Birk, A., 2017: The Pinax-model for accurate and efficient refraction correction of underwater cameras in flat-pane housings. Ocean Eng., 133, 9–22. <https://doi.org/10.1016/j.oceaneng.2017.01.029>.
- Mandlbürger, G., 2022: A review of active and passive optical methods in hydrography. Int. Hydrogr. Rev., 28, 8–52. <https://doi.org/10.58440/ihr-28-a15>.
- Menna, F., Nocerino, E., and Remondino, F.: Flat Versus Hemispherical Dome Ports in Underwater Photogrammetry, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W3, 481–487, <https://doi.org/10.5194/isprs-archives-XLII-2-W3-481-2017>, 2017
- Muhammad, F., Poerbandono, Sternberg, H., Djunarsjah, E., Abidin, H.Z., 2025: An appraisal of backscatter removal and refraction calibration models for improving the performance of vision-based mapping and navigation in shallow underwater environments. Intell. Syst. Appl., 25, 200476. <https://doi.org/10.1016/j.iswa.2025.200476>.
- Pedersen, M., Bengtson, S.H., Gade, R., Madsen, N., Moeslund, T.B., 2018: Camera calibration for underwater 3D reconstruction based on ray tracing using Snell's law. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work., 2018-June, 1491–1498. <https://doi.org/10.1109/CVPRW.2018.00190>.
- Rublee, E., Rabaud, V., Konolige, K., Bradski, G., 2014: ORB: An efficient alternative to SIFT or SURF. In: IEEE Int. Conf. Comput. Vis., Barcelona, Spain. <https://doi.org/10.1109/ICCV.2011.6126544>.
- Sarlin, P.-E., DeTone, D., Malisiewicz, T., Rabinovich, A., 2020: SuperGlue: Learning feature matching with graph neural networks. In: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 4938–4947.
- Sarlin, P.E., Cadena, C., Siegwart, R., Dymczyk, M., 2019: From coarse to fine: Robust hierarchical localization at large scale. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2019-June, 12708–12717. <https://doi.org/10.1109/CVPR.2019.01300>.
- Schönberger, J.L., Frahm, J.-M., 2016: Structure-from-Motion Revisited. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 4104–4113. <https://doi.org/10.1109/CVPR.2016.445>.
- She, M., Seegräber, F., Nakath, D., Köser, K., 2024: Refractive COLMAP: Refractive Structure-from-Motion Revisited. In: IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS). <https://doi.org/10.13140/RG.2.2.13570.08642>.
- Shortis, M., 2015: Calibration Techniques for Accurate Measurements by Underwater Camera Systems. Sensors, 15(12), 30810–30826. <https://doi.org/10.3390/s151229831>.
- Zhang, H., Chau, L.P., 2016: Removing backscatter to enhance the visibility of underwater object. Nanyang Technological University, Singapore.
- Zhou, J., Yang, T., Chu, W., Zhang, W., 2022: Underwater image restoration via backscatter pixel prior and color compensation. Eng. Appl. Artif. Intell., 111, 104785. <https://doi.org/10.1016/j.engappai.2022.104785>.