Zero-Shot Seafloor Sediment Microtopography Characterization Using Stereo from a Drifting Monocular Camera

Shahrokh Heidari[1,2⋆], Mihailo Azhar[3], Tegan Evans[2], Yani He[1], Stefano Schenone[2], Patrice Delmas[1], and Simon Thrush[2]

[1] IVSLab, The University of Auckland, Auckland 1010, New Zealand
[2] Institute of Marine Science, The University of Auckland, Auckland 1142, New Zealand
[3] Department of Ecoscience, Applied Marine Ecology and Modelling, Aarhus University, Roskilde, Denmark

Abstract. High-resolution characterization of seafloor sediment microtopography is essential for understanding benthic habitat structure, sedimentary processes, and ecological function. However, existing methods typically rely on core-based sampling or specialized 3D imaging systems, both of which are limited by cost, complexity, and scalability. In this study, we present a cost-effective, camera-based framework for quantitative sediment surface analysis using video footage from a drifting monocular underwater camera. The method leverages a zero-shot application of RAFT-Stereo to estimate dense disparity maps from sequential frames without requiring prior training on sediment data. After normalizing disparities, we apply surface detrending and extract statistical and morphological roughness features. Through a small-scale case study, we evaluate the method on two distinct sediment types, *Sand* and *Shell-Hash*, and demonstrate that the extracted features effectively capture surface complexity. Additionally, we assess the pipeline's consistency using overlapping samples acquired with different virtual stereo baselines, showing that key global features remain robust despite variations in camera motion. This framework offers a scalable, non-invasive solution for retrospective and in-situ sediment analysis in marine monitoring.

Keywords: Seafloor Microtopography · RAFT-Stereo · Zero-shot Depth Estimation.

1 Introduction

Understanding the seabed's biogeophysical characteristics is fundamental to effective marine ecosystem management, particularly in the context of the ongoing biodiversity crisis [15]. Detailed knowledge of biological (e.g., burrows and fecal mounds) and physical (e.g., sand ripples) structure and dynamics provides critical insights into broad-scale ecosystem processes and supports accurate assessments of how anthropogenic pressures impact benthic habitats, particularly those with high conservation value [19]. Additionally, recent studies have shown that the small-scale structural features on the sediment surface, shaped by the activity of resident fauna, reflect critical ecological processes such as organic matter degradation, nutrient cycling, and sediment-water exchange [31, 2, 30]. Traditional approaches to studying seabed sediments rely heavily on direct sampling techniques, including grabs, corers, trawls, and dredges [7]. In laboratory-based sediment core analyses, the standard workflow typically begins with a visual description of sedimentary structures, followed by a suite of physical and chemical assessments. These methods are often complemented by manual layer-by-layer examination to classify strata based on color and texture; a process that is labor-intensive, subject to observer bias, and prone to high uncertainty [16, 8]. While sediment cores provide essential insights into sediment composition, stratigraphy, and benthic habitat characteristics, their application is limited by spatial constraints, high operational costs, and logistical challenges in deep or turbid waters [6, 5]. In this context, the collection of seabed imagery through underwater photography and videography offers a complementary means to observe sediment surface features and microtopography, thereby enhancing core-based analyses with spatially extensive surface information. Visual data offer a direct, non-invasive means of assessing the underwater environment, enabling detailed observation of sediment types, surface structures, benthic fauna, and organism-sediment interactions. Importantly, imagery can capture ecologically relevant processes such as bioturbation, sediment resuspension, and microtopographic changes that are otherwise difficult to detect through core-based methods [4]. Furthermore, technological advancements in structure-from-motion (SfM) photogrammetry [30, 28, 9], computer vision and artificial intelligence [26, 29, 34, 37, 17] now allow for the quantitative extraction of microtopographic metrics, facilitating high-resolution spatial analysis of seafloor texture, structure and complexity. Advanced imaging techniques, including monocular

⋆ *Corresponding author*: shahrokh.heidari@auckland.ac.nz

and stereo camera systems, are increasingly employed to generate high-resolution 3D reconstructions of seabed sediments. These approaches enhance the accuracy of sediment characterization by enabling precise quantification of topographic attributes such as the variations in sediment surface roughness, thereby improving the assessment of sedimentary processes and morphological change. For instance, Cherisse et al. [10] introduced a Microtopographic Laser Scanning (MiLS) system to quantify fine-scale bottom roughness in fragile deep-sea habitats. Their method combined a downward-facing video camera and a single-point optical laser to capture bottom profiles at a resolution of 1–2 cm. While promising, MiLS requires specialized hardware, including a trolley-mounted camera-laser array, precise calibration, and controlled deployment conditions, limiting its operational flexibility and applicability in more dynamic environments. Similarly, Johnson-Roberson et al. [18] demonstrated a large-scale underwater 3D mapping system for archaeological site reconstruction in Greece. Their method employed both an autonomous underwater vehicle (AUV) and a diver-controlled stereo camera platform, integrated with a bundle-adjustment framework capable of handling over 400,000 stereo images to produce surface reconstructions at 2 $mm/pixel$ resolution across 26,600 m^2. Although this approach provides exceptional spatial detail, it involves significant logistical and computational demands, making it resource-intensive and potentially impractical for routine marine sediment monitoring. Recently, Schenone et al. [30] introduced a framework for quantifying sediment microtopography as a proxy for biodiversity and ecosystem functioning in subtidal soft-sediment habitats. Their method integrates SfM photogrammetry from diver-acquired video with ecological field sampling, benthic flux measurements, and macrofauna trait analysis. By calculating a set of quantitative surface metrics, the study demonstrated that microtopographic features strongly correlate with benthic biogeochemical fluxes, sediment properties, and macrofaunal diversity. However, the method also presents certain limitations. Its dependence on diver-based acquisition and manual deployment of incubation chambers restricts its spatial scalability and operational feasibility in deeper or less accessible environments. Furthermore, while SfM is effective at generating detailed 3D reconstructions, it requires stable lighting, consistent motion, and careful scene coverage to avoid reconstruction artifacts. The analysis pipeline also involves substantial computational effort [14].

Therefore, developing a cost-effective and scalable approach for generating high-resolution 3D reconstructions of sediment surfaces is critical for advancing sedimentary research and environmental monitoring. Importantly, an approach that balances affordability with precision can overcome the operational and financial limitations associated with traditional core-based sampling and diver-driven photogrammetry. This highlights the value of a cost-effective methodology that operates without the need for specialized instrumentation or rigid deployment protocols, making it particularly well-suited for retrospective analysis of video footage already collected during routine marine surveys. Such an approach leverages existing visual datasets, often acquired during benthic habitat assessments, ROV transects, or diver-based monitoring, allowing researchers to extract quantitative information on sediment microtopography and seabed structure without additional field effort or equipment investment.

Building on this premise, we aim to analyze seafloor sediment characteristics and fine-scale roughness, collectively referred to as sediment microtopography, using a low-cost, camera-based imaging solution. In this study, we present an approach that leverages optical imaging, photogrammetric principles, and artificial intelligence to extract microtopographic metrics from single-view video footage. Unlike conventional methods that require structured-light sensors, stereo rigs, or laser-based systems, our framework is uniquely capable of operating on pre-existing monocular video data, such as that collected during standard benthic transects or ROV deployments. As a result, the proposed method offers a scalable and accessible pathway for characterizing sediment morphology in both archival and future marine imaging datasets. Specifically, the proposed pipeline leverages two sequential frames from a drifting monocular camera to simulate a stereo-vision system. These temporally adjacent frames are first processed through stereo rectification to ensure Epipolar alignment. We then employ RAFT-Stereo [20], a deep learning-based stereo matching algorithm, in a zero-shot configuration (meaning the model is used without any fine-tuning or retraining on domain-specific sediment data) to estimate a dense disparity map between the frames. From this disparity output, we extract microtopographic metrics using a surface detrending and feature computation framework. To illustrate the practical utility of these features, we conduct a small-scale case study demonstrating how two example seabed types can be effectively distinguished based on their derived microtopographic profiles.

The rest of the paper is organized as follows. Section 2 presents the proposed methodology in detail. Section 3 provides the experimental results and discussion. Finally, Section 4 concludes the paper and outlines directions for future work.

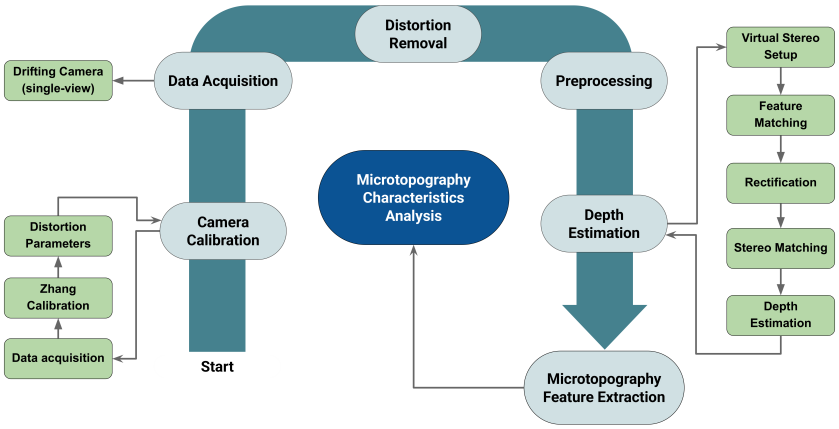

Fig. 1. The outline of the proposed microtopography characteristic analysis.

2 Methodology

In this section, we present a novel, cost-effective pipeline for characterizing sediment microtopography using video data acquired from a drifting single-view camera system. Although this study uses footage from a downward-facing drop camera system as a representative example, the proposed approach does not depend on a specific camera type. Any underwater camera capable of capturing continuous video while moving along a transect, whether towed, diver-operated, or vehicle-mounted, can be used, provided that the motion enables sufficient parallax for depth recovery. As illustrated in Figure 1, the framework includes six core components: camera calibration, data acquisition, distortion removal, preprocessing, depth estimation, and microtopographic feature extraction.

Camera calibration: it is a fundamental step in computer vision and photogrammetry, enabling the estimation of key parameters such as focal length, principal point offset, lens distortion, and the camera's spatial orientation. These intrinsic and extrinsic parameters are essential for accurate 3D reconstruction and distortion correction. A common approach uses a planar calibration target, typically a chessboard pattern, due to its regular grid structure and easily detectable corners. By capturing multiple images of the pattern from different viewpoints, corresponding 2D image points and 3D world points can be extracted and used to optimize the camera model via nonlinear least squares. In underwater imaging, calibration becomes even more critical due to the optical distortions introduced by the water medium. Refraction at the interface between the camera housing and the water column can produce nonlinear effects that differ from those observed in the air, often amplifying radial and tangential distortions. For this reason, calibration should be performed in the same optical medium in which the camera will operate to accurately capture the interaction between light and the lens-water interface.

Data Acquisition: Our pipeline relies on video data captured by a moving monocular camera, from which we extract two temporally spaced frames to simulate a stereo pair. Rather than using immediately consecutive frames, we intentionally skip a few frames between selections to ensure a sufficient baseline, that is, a larger camera displacement, to induce the parallax necessary for depth estimation. These frame pairs are treated as if they were captured by two spatially offset cameras, with the effective baseline corresponding to the camera's movement between the selected frames. This motion-based setup enables us to apply the geometric principles of stereo vision using a single monocular camera. To ensure accurate depth recovery and high-resolution reconstruction of sediment microtopography, the camera should remain as close as operationally feasible to the seabed. A low-altitude configuration enhances the visibility of fine-scale features and minimizes optical degradation caused by water column effects like scattering and turbidity.

Distortion Removal: Camera distortion refers to deviations between the captured image and the actual geometry of the scene, lens imperfections, optical aberrations, or environmental factors [1]. In underwater imaging, these distortions are further amplified by the water/air interface in the camera casing, water turbidity, and currents. Correcting for distortion is critical to ensure geometric consistency across frames, enabling accurate measurement, surface reconstruction, and visual interpretation. The intrinsic parameters and distortion coefficients estimated during the calibration phase correct the distortion.

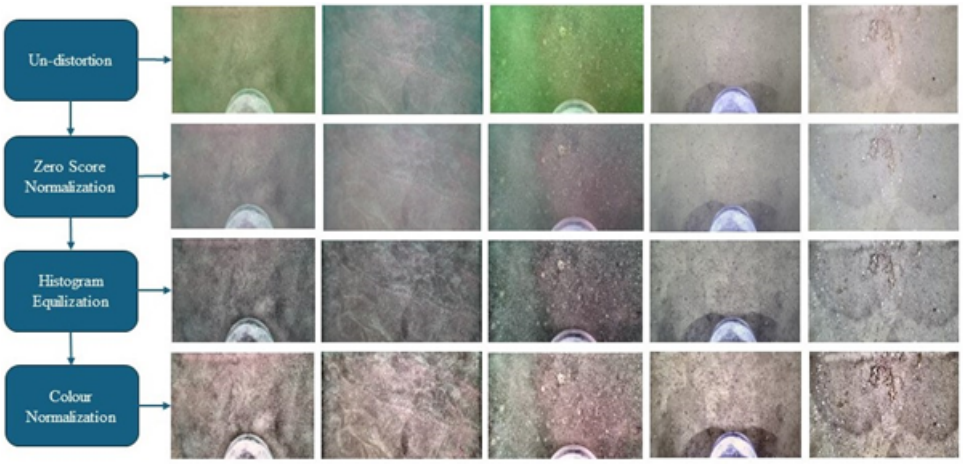

Fig. 2. Color Calibration examples based on the image processing approaches: *Z-Score Normalization*, *Contrast Enhancement*, and *Color Normalization*

Preprocessing: Underwater imaging is subject to significant color distortion due to the optical properties of water. Light penetrates the water column and interacts with suspended particles and dissolved substances, causing wavelength-dependent absorption and scattering. Longer wavelengths (e.g., red, orange, and yellow) are absorbed more rapidly, while shorter wavelengths (e.g., blue and green) penetrate deeper. This spectral imbalance results in the characteristic blue-green appearance of underwater environments and introduces substantial color degradation in captured imagery. To address this issue, color calibration is essential to restore a more accurate and consistent representation of the scene. Traditional color calibration methods involve imaging a color calibration target (e.g., a chessboard with known color patches) under controlled lighting. However, replicating true underwater spectral properties in a laboratory tank is challenging and may not produce valid calibration data. As a result, image-based postprocessing techniques are often favored for underwater color correction. In this study, we implemented a three-stage image enhancement pipeline for color calibration and normalization, designed to improve color consistency across video frames [22]: *Z-Score Normalization*, *Contrast Enhancement*, and *Color Normalization (Histogram Matching)*. This three-stage approach facilitates more reliable feature extraction and depth estimation in later stages of the pipeline by reducing noise and standardizing the visual properties of the input data, in particular the colour differences due to vignetting. Figure 2 illustrates raw video frames alongside their corresponding processed versions.

2.1 Depth Estimation

This section outlines an efficient method for estimating depth from a moving monocular camera by extracting two consecutive frames to simulate a stereo camera setup. In stereo vision, depth is inferred from two cameras capturing the scene from slightly different viewpoints [39]. Here, the same effect is achieved using a single moving camera, where the first frame is treated as the "left" view and the subsequent frame as the "right" view. The displacement between the two camera positions creates the necessary parallax for depth inference. While variations in camera speed can lead to inconsistencies in the virtual baseline, and thus affect depth accuracy, the goal of this method is not precise metric reconstruction. Instead, it focuses on capturing relative depth variation to characterize surface roughness and microtopographic features, allowing the identification of structural trends across the seafloor.

The disparity between corresponding pixels in the stereo pair is first estimated. Disparity refers to the horizontal shift in pixel position between the two rectified frames and is inversely related to depth: closer objects yield larger disparities. A key challenge in this process is stereo matching, which involves finding corresponding pixels between the two frames. To reduce its computational complexity, the images are first rectified, a transformation that aligns corresponding features along the same horizontal lines. This reduces the 2D search space to a 1D horizontal search, significantly simplifying and accelerating disparity computation. Given that our system uses a monocular moving camera and cannot access accurate extrinsic pose information between frames, we adopt an uncalibrated rectification approach. This involves detecting and matching features between two consecutive frames, estimating the Epipolar ge-

ometry, and rectifying the images accordingly. Our depth estimation pipeline thus consists of three main stages: feature matching, uncalibrated rectification, and stereo matching.

Feature matching involves the identification of corresponding features between two images. It consists of two main stages: feature detection and matching. We use Scale-Invariant Feature Transform (SIFT) for feature detection [35], and to establish correspondences between the detected SIFT features in both images, we employ a fast approximate nearest neighbor search using the FLANN (Fast Library for Approximate Nearest Neighbors) matcher with a KD-tree indexing algorithm [23].

For the stereo rectification part, we compute the fundamental matrix using the normalized eight-point algorithm [13]. This is followed by estimating two homography matrices, one for each view, that warp the images so that their epipolar lines become horizontal and aligned. This geometric transformation ensures that corresponding points lie on the same image rows, thus enabling efficient disparity computation. We adopt the approach of Fusiello et al. [11] for this part.

The next step is computing the disparity map based on the rectified pair of views. We use RAFT-Stereo [20], a deep learning-based stereo matching framework that builds upon the RAFT optical flow architecture to estimate dense disparity maps from rectified stereo image pairs. RAFT-Stereo employs a multi-level GRU-based update operator that performs iterative refinement of the disparity field using a lightweight 3D cost volume constructed via feature correlation across the same horizontal scanlines. A key strength of RAFT-Stereo lies in its zero-shot generalization capability. It achieves state-of-the-art results on real-world datasets (such as KITTI [12], and Middlebury [27]) despite being trained exclusively on synthetic data. This property makes it particularly suitable for our microtopography analysis of seafloor sediments, where annotated real-world training data is scarce.

The conversion from disparity to depth is a well-established process in stereo vision, based on the focal length and baseline (distance between the two camera poses). This relationship allows for the direct depth recovery from disparity when the baseline is known and consistent. However, in our monocular setup, where the disparity is inferred from temporally adjacent frames of a moving single-view camera, the effective baseline between frames varies due to uncontrolled camera motion. This introduces inconsistency in the depth scale across frame pairs, making it impractical to compute absolute metric depth values. Nevertheless, our analysis of seafloor sediment microtopography does not rely on absolute depth; rather, it focuses on relative surface variation and roughness characteristics. Applying a surface-detrending technique isolates fine-scale structural patterns while disregarding global elevation shifts. As such, disparity maps serve as a sufficient proxy for depth. Therefore, without loss of generality, we use disparity maps throughout our pipeline to capture microtopographic features effectively. Figure 3 illustrates the depth estimation process from a pair of sequential frames, highlighting each stage of our proposed pipeline.

2.2 Microtopography Feature Extraction

Given the variable and unknown baselines between sequential frames in our monocular setup, the resulting depth maps cannot be interpreted in absolute metric units. Consequently, our analysis relies on unit-invariant surface descriptors to characterize the microtopographic structure of the seafloor. We first normalize the depth map between zero and one. To isolate local topographic features from large-scale elevation trends, we adopt a surface-detrending approach based on the method proposed by Azhar et al. [2]. In this framework, the depth map is first reoriented so that the dominant surface normal aligns with the vertical axis (Z-axis), ensuring a consistent reference frame for detrending. This is achieved through two consecutive 3D rotations computed from principal component analysis (PCA) of the 3D surface points. Each pixel in the depth map is interpreted as a 3D point in Cartesian space, where its horizontal and vertical image indices correspond to the X and Y coordinates, and the depth value represents the Z coordinate. A first-order polynomial (i.e., a planar surface) of the form $z = a_0 + a_1x + a_2y$ is then fitted to the reoriented surface using least squares regression. The fitted plane represents the large-scale structure of the surface, which is then subtracted from the reoriented surface to produce a detrended map, denoted as Z_d. In this detrended image, local surface irregularities, such as depressions and mounds, are preserved as deviations from the plane, and are expressed as positive or negative values relative to the reference surface (see Figure 4). This output is the basis for subsequent microtopographic analysis using unit-independent statistical and morphological descriptors. Let N_u and N_v be the image horizontal, respectively vertical, dimensions in pixel unit, $Z_d(u, v)$ be the depth value at pixel location (u, v), and $\overline{Z_d}$ be the detrended surface mean. We adopt a set of surface metrics introduced by Azhar et al. [2] to characterize the microtopographic features of the sediment surface:

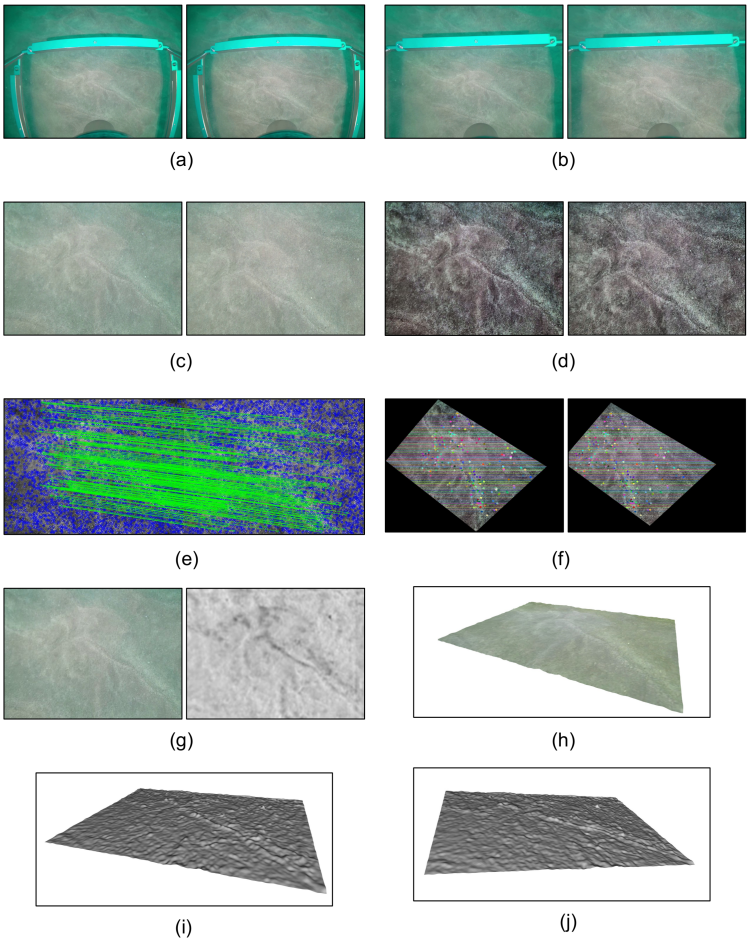

Fig. 3. Depth estimation on the sediment data examples. (a) two sequential video frames as a pair of stereo images, (b) undistorted video frames, (c), selected regions of interest (d) preprocessed frames, (e) feature matching, (f) stereo rectification results, (g) the first/left frame and the corresponding depth map computed using RAFT-Stereo. We also provide 3D meshes constructed from the depth map in (h)-(j) for visualization purposes.

$$S_a = \frac{1}{N_u N_v} \sum_{u=1}^{N_u} \sum_{v=1}^{N_v} |Z_d(u,v)|, \quad (1)$$

$$S_q = \sqrt{\frac{1}{N_u N_v} \sum_{u=1}^{N_u} \sum_{v=1}^{N_v} |Z_d^2(u,v)|}, \quad (2)$$

$$M_p = \frac{1}{N_u N_v} \sum_{u=1}^{N_u} \sum_{v=1}^{N_v} Z_d(u,v), \text{ where } Z_d(u,v) > 0, \quad (3)$$

$$\sigma_p = \sqrt{\frac{1}{N_u N_v} \sum_{u=1}^{N_u} \sum_{v=1}^{N_v} \left(Z_d(u,v) - \overline{Z_d}\right)^2}, Z_d(u,v) > 0, \quad (4)$$

$$M_t = \frac{1}{N_u N_v} \sum_{u=1}^{N_u} \sum_{v=1}^{N_v} Z_d(u,v), Z_d(u,v) < 0, \quad (5)$$

$$\sigma_t = \sqrt{\frac{1}{N_u N_v} \sum_{u=1}^{N_u} \sum_{v=1}^{N_v} \left(Z_d(u,v) - \overline{Z_d}\right)^2}, Z_d(u,v) < 0, \quad (6)$$

$$skew = \frac{1}{N_u N_v S_q^3} \sum_{u=1}^{N_u} \sum_{v=1}^{N_v} \left(Z_d(u,v)\right)^3, \quad (7)$$

$$kurtosis = \frac{1}{N_u N_v S_q^4} \sum_{u=1}^{N_u} \sum_{v=1}^{N_v} \left(Z_d(u,v)\right)^4, \quad (8)$$

where S_a is the mean of absolute deviations from the detrended surface, S_q is the root-mean-squared value of vertical deviations from the detrended surface, M_p is the mean height of positive surface pixel values, σ_p is the standard deviation of positive surface pixel values, M_t is the mean height of negative surface pixel values, σ_t the standard deviation of negative surface pixel values, *skew* quantifies the asymmetry of the surface height distribution, indicating whether mounds or depressions dominate the sediment structure, and *kurtosis* measures the peakedness of the distribution, reflecting the prevalence of extreme features. In addition to these, we extract microtopographic descriptors based on the Abbott-Firestone curve, including s_k, s_{pk}, and s_{vk}, which respectively quantify the core roughness, reduced peak height, and reduced valley depth (see Figure 5). We also use the ratios s_{pk}/s_k and s_{vk}/s_k, which capture the

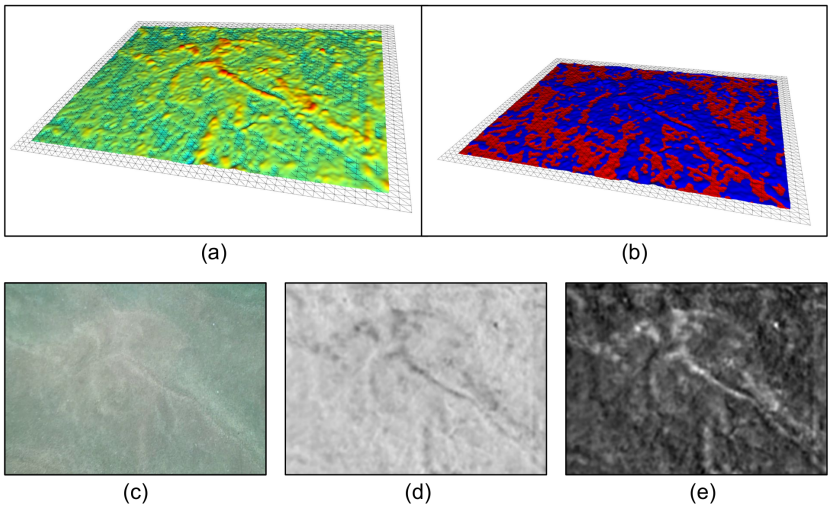

Fig. 4. Illustration of the depth map detrending process for microtopography analysis. The 3D surfaces have been scaled for visualization purposes. (a) The fitted plane is estimated from the depth map and visualized with a heat map to highlight large-scale surface trends. (b) The same fitted plane with regions above the plane is shown in blue and those below in red, emphasizing local deviations. (c) The original left image frame was used as input. (d) Estimated depth map corresponding to the input frame pair. (e) Final detrended depth map, isolating fine-scale surface roughness after removal of the planar trend.

relative prominence of peaks and valleys with respect to the core surface.

To complement these spatial-domain surface features, we additionally incorporate frequency-domain

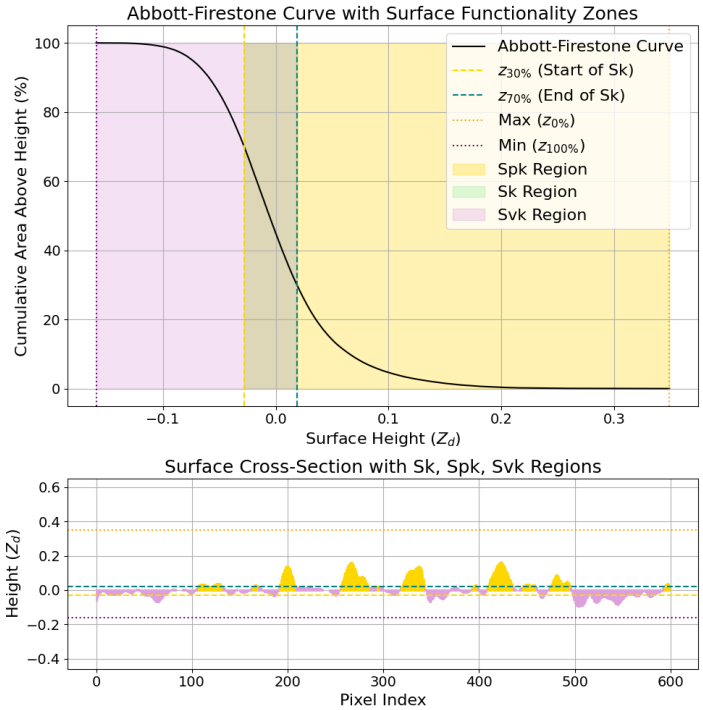

Fig. 5. Abbott-Firestone curve and surface functional zones on an example sediment surface. This plot represents the cumulative surface area distribution above each height value for a detrended sediment surface. The X-axis shows surface elevation, while the Y-axis indicates the percentage of the surface lying above that elevation. The curve is segmented into functional zones, valley S_{vk}, core roughness S_k, and peak S_{pk}, defined by standard thresholds. The decreasing trend reflects how surface features are distributed vertically, allowing quantitative characterization of microtopographic roughness.

features derived from the normalized depth map's fast fourier transform (FFT) and 2D power spectral density (PSD). The PSD quantifies how surface elevation's variance (or "power") is distributed across different spatial frequencies, effectively revealing the dominant structural scales and periodic patterns present in the microtopography. In a 2D PSD map, low-frequency components capture broad, gently varying features, while high-frequency components reflect fine-scale textures or abrupt transitions [3].

3 Experimental Results and Discussion

The experiments were conducted on a computing system with an NVIDIA GeForce RTX 4090 GPU and an Intel(R) Core(TM) i7-7700K CPU. We used the OpenCV library [24] for the camera calibration, distortion removal, preprocessing, feature-matching, and stereo rectification steps. The feature matching part involved the OpenCV implementation of SIFT, followed by the FLANN-based matcher, which was configured with five trees and 50 search checks to balance speed and accuracy. Regarding the stereo-matching part, we used the official GitHub repository of RAFT-Stereo [21] pre-trained on ETH3D stereo dataset [33]. To illustrate the ecological relevance of the proposed microtopographic features, we conducted a small-scale case study on *Sand* and *Shell-Hash* seabed types. Specifically, we selected two representative locations around Te Hauturu-o-Toi/Little Barrier Island, each known to be dominated by either *Sand* or *Shell-Hash* substrates (see Figure 7). We used an existing video dataset captured by a Bay Dynamics MK2 drop camera system. We employed Zhang's calibration method [38] to obtain intrinsic parameters from the camera. To simulate realistic underwater conditions, we calibrated the camera at the Goat Island Marine Discovery Centre, Auckland, New Zealand, using a seawater-filled tank. This controlled environment allowed us to avoid the practical difficulties of deploying calibration targets on the seafloor while maintaining calibration fidelity. Figure 6 illustrates the drop camera setup and sample calibration imagery, while Table 1 summarizes the estimated calibration parameters, respectively, which were subsequently used for distortion removal.

Fig. 6. Calibration setup used to simulate an underwater environment. The first row shows the drop camera system and the water-filled tank used for the calibration. The remaining images illustrate sample calibration data captured using the drop camera.

Table 1. Estimated camera calibration parameters using Zhang method [38]. The units are in *pixels*.

$\mathbf{f_x}$	$\mathbf{f_y}$	$\mathbf{c_x}$	$\mathbf{c_x}$	$\mathbf{k_1}$	$\mathbf{k_2}$	$\mathbf{p_1}$	$\mathbf{p_2}$	$\mathbf{k_3}$	**error**
2820.29	2823.79	2057.93	1525.15	−0.42	0.24	−0.0007	0.0004	−0.087	0.18

From each site, three stereo pairs of frames were extracted from underwater transects captured using the same monocular moving camera. After identifying regions of interest, the selected frame pairs were

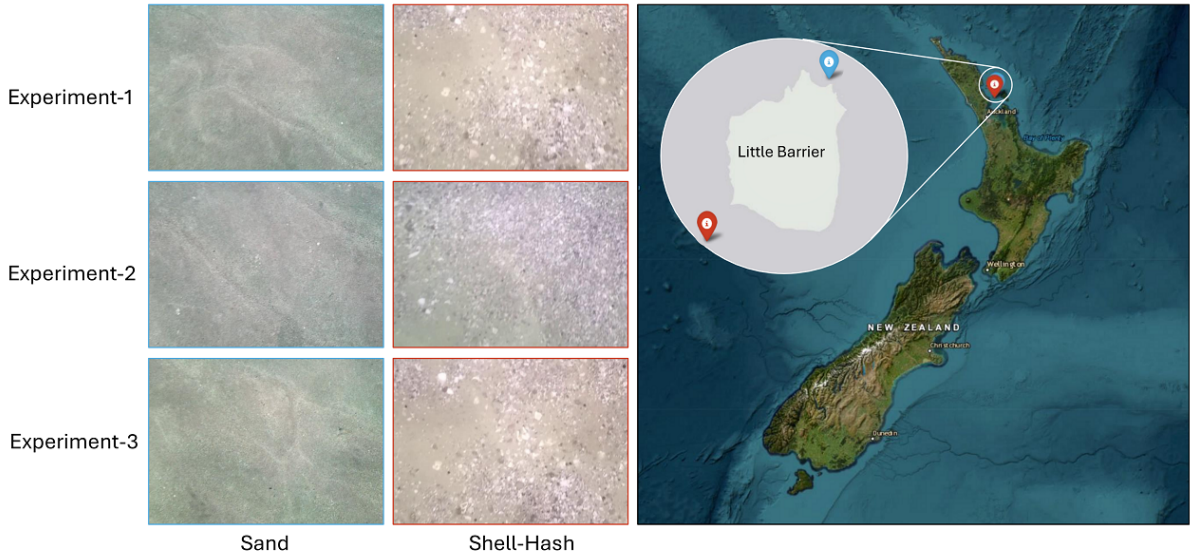

Fig. 7. Case study locations and sample frames for *Sand* and *Shell-Hash* seabed types. The blue marker corresponds to *Sand*-dominated substrates, while the red marker indicates *Shell-Hash*-dominated sites. The figure also shows representative undistorted left frames extracted from underwater transects at each site using the monocular moving camera system.

resized to 640×480 pixels for input to the RAFT-Stereo model, a resolution we found to offer a reasonable trade-off between processing speed and depth estimation performance. We applied our microtopography characterization pipeline to each sample to extract corresponding depth maps from sequential image pairs. Since the stereo matching component operates in a zero-shot manner, meaning it was not fine-tuned or trained on our specific data or sediment types, we ensured robustness by constraining the input to well-overlapping regions (nearly 90% between left and right frames). To achieve this, we cropped a square region centered in each depth map, maximizing the likelihood of consistent visual correspondence between frames. Microtopographic features were then extracted from these central regions for each sample. For clarity, from top to bottom, we refer to the paired samples in Figure 7 as Experiment-1, Experiment-2, and Experiment-3.

We first report the stereo rectification error for each sample in the experiments by measuring the average vertical disparity between matched features before and after rectification. Before the rectification, the vertical error is computed as the mean absolute difference in the y-coordinates of corresponding matched features across the stereo pair. The same features are projected using the estimated Homographies following rectification, and the vertical disparities are recalculated. A successful rectification substantially reduces this vertical alignment error, indicating improved geometric consistency between the stereo images (see Table 2).

Table 2. Rectification error (before/after) for each sediment surface in the experiments. The units are in *pixels*

Experiments	**Sediment Type**	**Rect-Error (Before)**	**Rect-Error (After)**
Experiment-1	*Sand*	23.71	0.36
	Shell-Hash	43.44	0.40
Experiment-2	*Sand*	14.22	0.41
	Shell-Hash	34.89	0.42
Experiment-3	*Sand*	24.30	0.38
	Shell-Hash	41.67	0.38

In Figure 7, we highlight the *Shell-Hash* samples from Experiment-1 and Experiment-3, which feature overlapping seafloor regions in their respective left images. The associated paired images, however, were

captured at different spatial intervals, resulting in varying virtual stereo baselines. These samples were selected to assess whether our depth-based surface analysis remains consistent across different stereo configurations for the same region. While both samples contain overlapping content, they also include non-overlapping areas, which introduces some variability. Nevertheless, we expect global topographic patterns to remain consistent across the region (as shown in Figure 8). In particular, features such as Sa, S_q, *skew*, and S_k should be robust, as they capture broad-scale structural properties of the surface. In contrast, features such as *kurtosis*, S_{pk}, S_{vk}, S_{pk}/S_k, and S_{vk}/S_k are more sensitive to local surface fluctuations and statistical extremes. These features are influenced by localized noise, minor differences in feature localization, and the presence or absence of small peaks and depressions, especially in non-overlapping regions, making them less reliable for cross-sample comparison under these conditions.

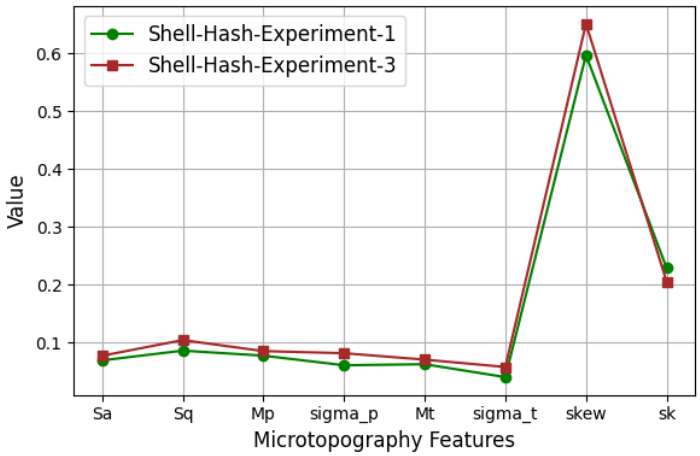

Fig. 8. Selected global roughness features extracted from two *Shell-Hash* samples (Experiment-1 and Experiment-3) with overlapping seafloor regions but different virtual stereo baselines.

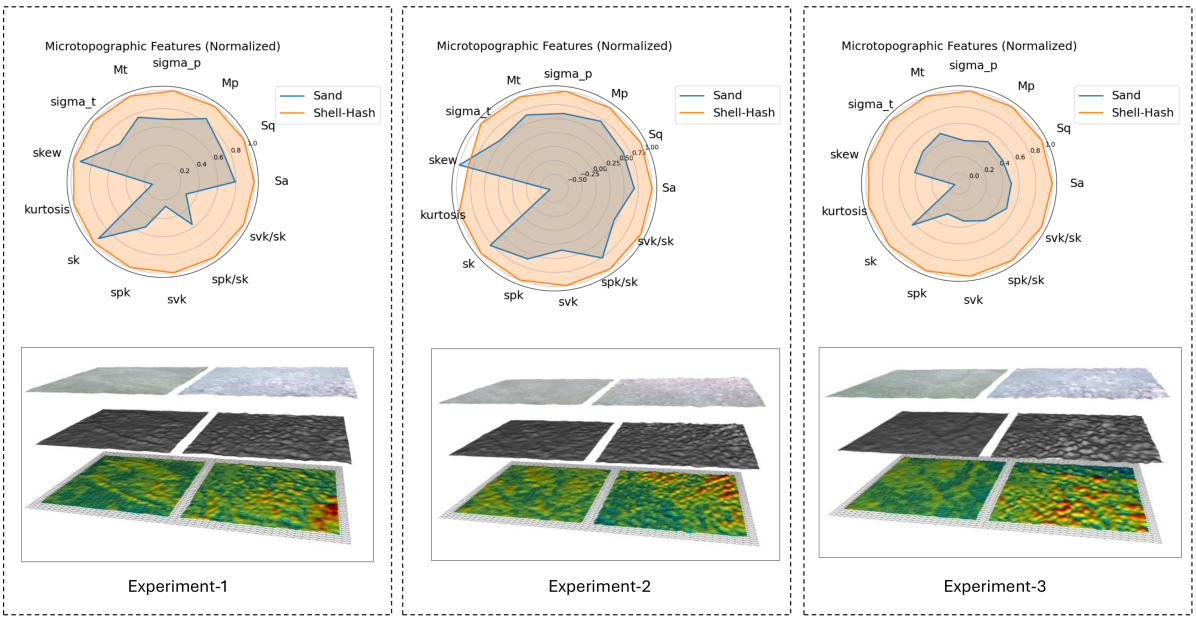

Fig. 9. Normalized microtopographic feature comparison across sample pairs. Radar plots comparing normalized microtopographic features between *Sand* and *Shell-hash* samples for three experimental pairs. Each plot corresponds to a sample pair from the same row in Figure 7, referred to as Experiment-1, Experiment-2, and Experiment-3 from top to bottom.

Figure 9 compares the two sediment types using their corresponding normalized microtopographic features visualized as radar plots. Each radar plot corresponds to a pair of samples taken from the same row in Figure 7. Across all three experimental comparisons, the extracted microtopographic features consistently exhibit higher values for *Shell-Hash* surfaces than for *Sand*, reflecting their inherently more complex and irregular structure. S_a and S_q quantify overall surface deviation, and their elevated values in *Shell-Hash* samples indicate greater textural variability and roughness, as expected from fragmented biogenic material compared to smoother sandy substrates. Similarly, M_p and M_t heights are higher in *Shell-Hash*, capturing the presence of more pronounced local maxima and minima in surface elevation. The corresponding standard deviations σ_p and σ_t further confirm greater variability in peak and trough amplitudes for *Shell-Hash*. The skewness metric, which indicates the asymmetry of the depth distribution, is generally higher for *Shell-Hash*, though one *Sand* sample exhibited greater skew. This may result from isolated protrusions or depressions in the *Sand* surface that skew the distribution more than a uniformly irregular *Shell-Hash* texture. Meanwhile, kurtosis, a measure of peakedness, is consistently higher for *Shell-Hash*, reflecting sharper elevation transitions and more extreme surface variations. The Abbott-Firestone metrics are also elevated in *Shell-Hash* samples, further supporting the presence of more extreme topographic features. Collectively, these results highlight the discriminatory power of the proposed microtopographic characterization pipeline and validate the ecological relevance of the extracted features in capturing sedimentary complexity.

Figure 10 presents the log-scale 2D PSD maps and their corresponding radial plots for each sediment surface. The 2D PSD maps illustrate how surface roughness is distributed across different spatial frequencies and directions. When the plots appear circular and evenly distributed, it suggests the surface is isotropic, meaning it has similar roughness in all directions. Bright centres in these plots indicate strong low-frequency content, corresponding to larger, smoother surface features. In contrast, brighter edges suggest higher-frequency content, which reflects finer, more detailed textures. The radial plots simplify this information into a one-dimensional curve, showing how the power changes with frequency. In these plots, low frequencies represent large-scale, gradual changes on the surface (coarser features), while high frequencies correspond to small, detailed variations (finer textures). High power at low frequencies means the surface is dominated by broad, smooth structures. In contrast, high power at high frequencies indicates more intricate, fine-scale roughness. Together, these visualizations allow for a detailed comparison of surface texture and complexity across different sediment types, highlighting how roughness varies at different spatial scales [36, 25]. These plots also enable the analysis of specific quantitative features that

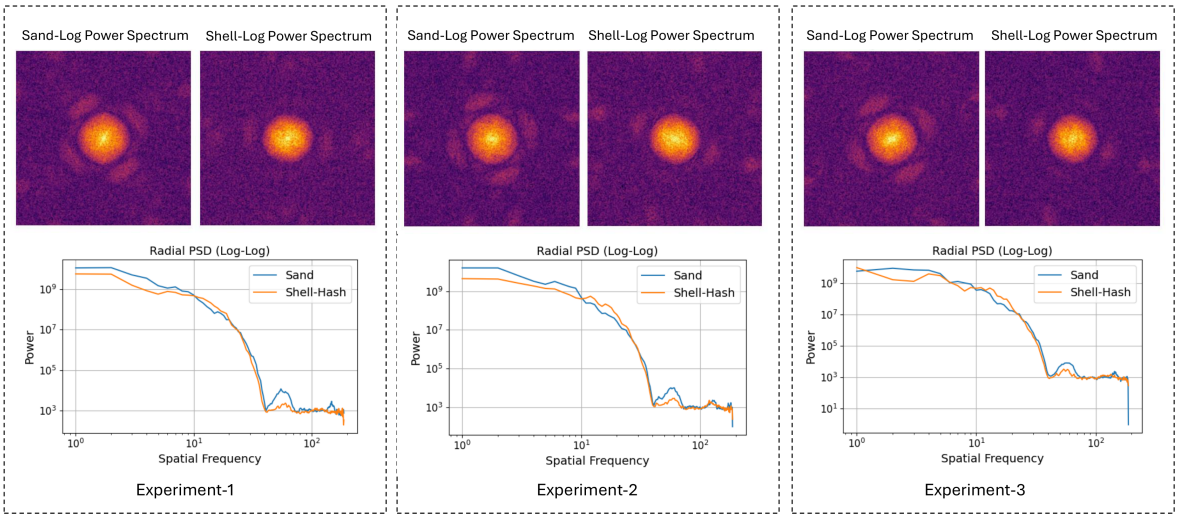

Fig. 10. Log-scale 2D PSD maps and corresponding radial plots for each sediment surface in our experiments. The 2D PSD maps show how surface roughness varies across spatial frequencies and directions. The Radial plots summarize this in 1D, with higher low-frequency power pointing to large-scale features and high-frequency power indicating fine-scale roughness.

are useful for comparing sediment surface types. For example, one could compute the spectral slope (describing how power decreases with frequency), the fractal dimension (quantifying surface complexity

and self-similarity across scales), the spectral centroid (indicating the average spatial frequency), and the high-frequency ratio (capturing the proportion of fine-scale content). Such features can help characterize surface texture and complexity more systematically.

Although the proposed zero-shot sediment microtopography framework is cost-effective and compatible with any underwater camera or existing video dataset, it has specific limitations. The method relies on a moving camera to generate parallax, requiring footage where the camera drifts alongside the sediment surface. We observed that the camera must remain close enough to the seafloor to capture fine-scale details effectively. Moreover, the success of feature matching and stereo rectification is highly dependent on the presence of distinctive texture features in the sediment. These factors can limit the robustness and generalizability of the method, especially in low-texture or high-turbidity environments. In ideal conditions, the proposed pipeline can produce distinct microtopographic signatures valuable for differentiating between sediment types. These features offer a scalable, non-invasive alternative to traditional sampling and can enhance existing classification frameworks. When integrated with other marine data sources (e.g., macrofaunal community data, measures of ecosystem function and biogeochemical properties), the microtopographic information can support a more holistic understanding of seafloor characteristics [29, 32, 30].

4 Conclusion

This study introduces a novel framework for seafloor sediment microtopography characterization that leverages stereo depth estimation from temporally adjacent frames of a drifting monocular camera. The method effectively isolates fine-scale surface variation from uncontrolled camera motion and unknown baselines by applying RAFT-Stereo in a zero-shot configuration and incorporating disparity normalization, uncalibrated rectification, and surface detrending. We validated the pipeline through experiments on two sediment types using field-acquired video data. We demonstrated its ability to extract ecologically meaningful microtopographic features from the sediment surfaces while maintaining consistent results across variable virtual stereo baselines within the same region. The approach addresses key limitations of traditional sampling and photogrammetric methods by eliminating the need for specialized hardware such as stereo camera rigs, laser scanners, structured-light systems, and reliance on annotated training datasets. It offers a scalable, non-invasive solution for extracting 3D surface information from existing or routinely collected underwater imagery. Future research will improve robustness in low-texture environments, integrate additional environmental data sources, and extend the pipeline for seafloor sediment classification and temporal change detection.

Acknowledgment

This research was funded by the Oceans of Change project and MBIE UOAX2307, and was supported by Our Seas Our Future. We thank J. Hillman, E. Ferretti, A. West, S. Thomas, G Cunningham, S. Ladewig, X. Yu and B. Doak for their assistance with data collection.

References

1. Ahmed, M., Farag, A.: Nonmetric calibration of camera lens distortion: differential methods and robust estimation. IEEE Transactions on image processing **14**(8), 1215–1230 (2005)
2. Azhar, M., Hillman, J.R., Gee, T., Schenone, S., van der Mark, W., Thrush, S.F., Delmas, P.: An rgb-d framework for capturing soft-sediment microtopography. Methods in Ecology and Evolution **13**(8), 1730–1745 (2022)
3. Briggs, K.B., Lyons, A.P., Pouliquen, E., Mayer, L.A., Richardson, M.D.: Seafloor roughness, sediment grain size, and temporal stability. Underwater Acoustic Measurements: Technologies and Results p. 8 (2005)
4. Buhl-Mortensen, L., Buhl-Mortensen, P., Dolan, M., Gonzalez-Mirelis, G.: Habitat mapping as a tool for conservation and sustainable use of marine resources: Some perspectives from the mareano programme, norway. Journal of Sea Research **100**, 46–61 (2015)
5. Cooper, K.M., Bolam, S.G., Downie, A.L., Barry, J.: Biological-based habitat classification approaches promote cost-efficient monitoring: An example using seabed assemblages. Journal of Applied Ecology **56**(5), 1085–1098 (2019)
6. Copeland, A., Edinger, E., Devillers, R., Bell, T., LeBlanc, P., Wroblewski, J.: Marine habitat mapping in support of marine protected area management in a subarctic fjord: Gilbert bay, labrador, canada. Journal of Coastal Conservation **17**, 225–237 (2013)

7. Cui, X., Yang, F., Wang, X., Ai, B., Luo, Y., Ma, D.: Deep learning model for seabed sediment classification based on fuzzy ranking feature optimization. Marine Geology **432**, 106390 (2021)
8. Di Martino, A., Carlini, G., Castellani, G., Remondini, D., Amorosi, A.: Sediment core analysis using artificial intelligence. Scientific Reports **13**(1), 20409 (2023)
9. Douglas, T.J., Coops, N.C., Drever, M.C.: Uav-acquired imagery with photogrammetry provides accurate measures of mudflat elevation gradients and microtopography for investigating microphytobenthos patterning. Science of Remote Sensing **7**, 100089 (2023)
10. Du Preez, C., Tunnicliffe, V.: A new video survey method of microtopographic laser scanning (mils) to measure small-scale seafloor bottom roughness. Limnology and Oceanography: Methods **10**(11), 899–909 (2012)
11. Fusiello, A., Trucco, E., Verri, A.: A compact algorithm for rectification of stereo pairs. Machine vision and applications **12**, 16–22 (2000)
12. Geiger, A., Lenz, P., Urtasun, R.: Are we ready for autonomous driving? the kitti vision benchmark suite. In: 2012 IEEE conference on computer vision and pattern recognition. pp. 3354–3361. IEEE (2012)
13. Hartley, R.I.: In defense of the eight-point algorithm. IEEE Transactions on pattern analysis and machine intelligence **19**(6), 580–593 (1997)
14. Hu, K., Wang, T., Shen, C., Weng, C., Zhou, F., Xia, M., Weng, L.: Overview of underwater 3d reconstruction technology based on optical images. Journal of Marine Science and Engineering **11**(5), 949 (2023)
15. Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES): Global assessment report on biodiversity and ecosystem services (2019). https://doi.org/10.5281/zenodo.6417333, https://zenodo.org/record/6417333, accessed: 2025-04-03
16. Jacq, K., Rapuc, W., Benoit, A., Coquin, D., Fanget, B., Perrette, Y., Sabatier, P., Wilhelm, B., Debret, M., Pignol, C., et al.: Sedimentary structure discrimination with hyperspectral imaging in sediment cores [data set]. Zenodo (2021)
17. Jahanbakht, M., Xiang, W., Azghadi, M.R.: Sediment prediction in the great barrier reef using vision transformer with finite element analysis. Neural Networks **152**, 311–321 (2022)
18. Johnson-Roberson, M., Bryson, M., Friedman, A., Pizarro, O., Troni, G., Ozog, P., Henderson, J.C.: High-resolution underwater robotic vision-based mapping and three-dimensional reconstruction for archaeology. Journal of Field Robotics **34**(4), 625–643 (2017)
19. Kenny, A.J., Cato, I., Desprez, M., Fader, G., Schüttenhelm, R., Side, J.: An overview of seabed-mapping technologies in the context of marine habitat classification. ICES Journal of Marine Science **60**(2), 411–418 (2003)
20. Lipson, L., Teed, Z., Deng, J.: Raft-stereo: Multilevel recurrent field transforms for stereo matching. In: 2021 International Conference on 3D Vision (3DV). pp. 218–227. IEEE (2021)
21. Lipson, L., Teed, Z., Deng, J.: Raft-stereo: Multilevel recurrent field transforms for stereo matching. https://github.com/princeton-vl/RAFT-Stereo (2021), accessed: 2025-02-15
22. Mbani, B., Greinert, J.: Analysis-ready optical underwater images of manganese-nodule covered seafloor of the clarion-clipperton zone. Scientific Data **10**(1), 316 (2023)
23. Muja, M., Lowe, D.G.: Fast approximate nearest neighbors with automatic algorithm configuration. VISAPP (1) **2**(331-340), 2 (2009)
24. OpenCV contributors: Opencv: Open source computer vision library. https://opencv.org/ (2024), accessed: 2024-04-01
25. Rodriguez, N., Gontard, L., Ma, C., Xu, R., Persson, B.: On how to determine surface roughness power spectra. Tribology Letters **73**(1), 18 (2025)
26. Saleh, M., Rabah, M.: Seabed sub-bottom sediment classification using parametric sub-bottom profiler. NRIAG Journal of Astronomy and Geophysics **5**(1), 87–95 (2016)
27. Scharstein, D., Hirschmüller, H., Kitajima, Y., Krathwohl, G., Nešić, N., Wang, X., Westling, P.: High-resolution stereo datasets with subpixel-accurate ground truth. In: Pattern Recognition: 36th German Conference, GCPR 2014, Münster, Germany, September 2-5, 2014, Proceedings 36. pp. 31–42. Springer (2014)
28. Schenone, S., Azhar, M., Delmas, P., Thrush, S.F.: Towards time and cost-efficient habitat assessment: challenges and opportunities for benthic ecology and management. Aquatic Conservation: Marine and Freshwater Ecosystems **33**(12), 1603–1614 (2023)
29. Schenone, S., Azhar, M., Ramírez, C.A.V., Strozzi, A.G., Delmas, P., Thrush, S.F.: Mapping the delivery of ecological functions combining field collected data and unmanned aerial vehicles (uavs). Ecosystems pp. 1–12 (2021)
30. Schenone, S., Hewitt, J.E., Hillman, J., Gladstone-Gallagher, R., Gammal, J., Pilditch, C., Lohrer, A.M., Ferretti, E., Azhar, M., Delmas, P., et al.: Seafloor sediment microtopography as a surrogate for biodiversity and ecosystem functioning. Ecological Applications **35**(1), e3069 (2025)
31. Schenone, S., Thrush, S.F.: Unraveling ecosystem functioning in intertidal soft sediments: the role of density-driven interactions. Scientific Reports **10**(1), 11909 (2020)
32. Schenone, S., Thrush, S.F.: Scaling-up ecosystem functions of coastal heterogeneous sediments: testing practices using high resolution data. Landscape Ecology **37**(6), 1603–1614 (2022)
33. Schops, T., Schonberger, J.L., Galliani, S., Sattler, T., Schindler, K., Pollefeys, M., Geiger, A.: A multi-view stereo benchmark with high-resolution images and multi-camera videos. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3260–3269 (2017)

34. Wan, J., Qin, Z., Cui, X., Yang, F., Yasir, M., Ma, B., Liu, X.: Mbes seabed sediment classification based on a decision fusion method using deep learning model. Remote Sensing **14**(15), 3708 (2022)
35. Wu, J., Cui, Z., Sheng, V.S., Zhao, P., Su, D., Gong, S.: A comparative study of sift and its variants. Measurement science review **13**(3), 122 (2013)
36. Yang, H., Baudet, B.A., Yao, T.: Characterization of the surface roughness of sand particles using an advanced fractal approach. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **472**(2194), 20160524 (2016)
37. Zhang, Q., Zhao, J., Li, S., Zhang, H.: Seabed sediment classification using spatial statistical characteristics. Journal of Marine Science and Engineering **10**(5), 691 (2022)
38. Zhang, Z.: A flexible new technique for camera calibration. IEEE Transactions on pattern analysis and machine intelligence **22**(11), 1330–1334 (2002)
39. Zhou, K., Meng, X., Cheng, B.: Review of stereo matching algorithms based on deep learning. Computational intelligence and neuroscience **2020**(1), 8562323 (2020)