

# A COMBINED APPROACH FOR LONG-TERM MONITORING OF BENTHOS IN ANTARCTICA WITH UNDERWATER PHOTOGRAMMETRY AND IMAGE UNDERSTANDING

F. Menna <sup>1\*</sup>, E. Nocerino <sup>2</sup>, S. Malek <sup>1</sup>, F. Remondino <sup>1</sup>, S. Schiaparelli <sup>3,4</sup>

<sup>1</sup> 3D Optical Metrology (3DOM) unit, Bruno Kessler Foundation (FBK), Trento, Italy – (fmenna, smalek, remondino)@fbk.eu

<sup>2</sup> Dipartimento di Scienze Umanistiche e Sociali, Università degli Studi di Sassari, Sassari, Italy - enocerino@uniss.it

<sup>3</sup> DISTAV, University of Genoa, Genoa, Italy - stefano.schiaparelli@unige.it

<sup>4</sup> Italian national Antarctic Museum (MNA, Section of Genoa), Genoa, Italy

## Commission II, WG II/9

**KEY WORDS:** Underwater photogrammetry, change detection, long-term monitoring, antarctica, image segmentation, CNN

### ABSTRACT:

Long-term monitoring projects are becoming more than ever crucial in assessing the effects of climate change on marine communities, especially in Antarctica, where these changes are expected to be particularly dramatic. Detailed studies of the Antarctic benthos are in fact particularly important for a better understanding of benthos dynamics and potential climate-driven shifts. Here, due to the extreme fragility of benthic communities, non-destructive techniques represent the best solution in long-term monitoring programs. In this paper we report new results from 2017, 2018, 2019 photogrammetric campaigns within the Italian National Antarctic Research Program (PNRA). A new protocol of data acquisition and multi-temporal processing that provides co-registered 3D point clouds between the three years without control points nor direct georeferencing methods is presented. This is achieved by adding a level of image understanding leveraging semantic segmentation with convolutional neural network (CNN) of the benthic features. Slow growing (estimated less than a mm per year) organisms, such as Corallinales (Rhodophyta algae), represent a natural stable pattern, leveraged to automatically orient in the same reference system the photogrammetric surveys of different epochs. This approach is also proved to be effective in improving the orientation of adjacent strips acquired within the same campaign. Within the paper an in depth analysis of the achieved results shows the effectiveness of the implemented procedure.

## 1. INTRODUCTION

### 1.1 Time series for Long Term Ecological Research

The effects of climate change are becoming particularly dramatic in Antarctica and long-term monitoring projects are more than ever crucial to assessing its impact on habitat and species, both terrestrial and marine. Although field conditions are extreme and cost prohibitive, several Long Term Ecological Research (LTER, Callahan 1984; Strayer et al. 1986) projects have been implemented over the years. Some examples include: (i) coastal oceanography (Palmer station LTER program, <https://pal.lternet.edu/>, Smith et al. 1995); (ii) environmental impacts of research bases and human presence (Conlan et al. 2010; Stark et al. 2006); (iii) time-series data about pinniped populations, krill and plankton (Hosie et al., 2014); (iv) the frequency of ice scouring events on benthic biota (Deregibus et al. 2017). In order to address the lack of long-term studies on the benthos dynamics in Antarctic marine ecosystems (Smith et al. 1995), the SCAR (Scientific Committee on Antarctic Research) established an *ad hoc* working group, the Antarctic Near-Shore And Terrestrial Observation System (ANTOS) (<https://www.scar.org/science/antos/home/>), to coordinate this research and pinpoint new monitoring techniques tailored to the peculiarities of Antarctic environments. Detailed studies of the Antarctic benthos are particularly important for a better understanding of a number of processes, such as (i) the growth performance in slow-growing and long-living key species, (ii) the impacts and disturbance of episodic events (e.g. iceberg-scouring), (iii) the inter-annual variability in food delivery, and

(iv) the shifts in community structure not due to stochastic events.

### 1.2 Related works: non-destructive techniques for benthic communities

The measurement techniques employed in long term monitoring programs focusing on benthic communities are differentiated into ‘extractive’ or ‘destructive’ and ‘non-destructive’ methodologies, the latter being preferred as they preserve the ecosystem (Solan et al. 2003). In the last years, several non-destructive approaches have been adapted for the study of benthic dynamics and spatial patterns. These include: (i) underwater optical recording systems, including both digital video and still photography (e.g. Chennu et al. 2017; Peirano et al. 2016), (ii) photogrammetry for 3D surveying and modelling of the seafloor and its features (e.g. Nocerino et al. 2020; Pizarro et al. 2017) and (iii) AI (artificial intelligence)-assisted software for benthic imagery automatic analysis (e.g. Pavoni et al. 2021; Trygonis and Sini 2012). Photographic and/or video sampling is of great value as it permanently records the current situation, thus enabling comprehensive image analyses, reducing the time spent underwater and the need of experienced divers in species identification (Parravicini et al. 2009), although metric quantitative analyses are of low accuracy and limited to small areas (e.g. photo quadrats of 1m<sup>2</sup>). On the other hand, underwater photogrammetry is increasingly being used by marine ecologists because of its ability to produce accurate, spatially detailed, non-destructive measurements of benthic communities, coupled with affordability and ease of use (Pavoni

\* Corresponding author

et al., 2021; Nocerino et al., 2020; Piazza et al., 2018; 2019 and 2020; Pizzarro et al., 2017). However, independent quality control, rigorous imaging system set-up, optimal geometry design and a strict modeling of the imaging process are essential to achieve a high degree of measurable accuracy and resolution. If a proper photogrammetric approach that enables the formal description of the propagation of measurement error and modeling uncertainties is not undertaken, statements regarding the statistical significance of the measured changes are limited. In Nocerino et al. (2020) the multi-temporal 3D monitoring via underwater photogrammetry is described in the case of coral reef to better understand factors that mediate coral community structure and function. Accurate reference networks of targeted points are required for environmental change detection and monitoring as already demonstrated by first application of the technique in Antarctica (Piazza et al., 2018; 2019 and 2020). They are crucial when the variations to be measured are in the range of a few centimeters per year, typical of highly dynamic environments such as oceanic coral reefs, where 3D landscape elements are continuously changing over time by growing or shrinking or fragmenting or when sand can be deposited on organisms.

### 1.3 Towards fully automatic long-term monitoring in Antarctica using underwater photogrammetry

A network of stable targets on the seabed as described in Nocerino et al. (2020) provides the most accurate and reliable method for change detection using photogrammetry. Spatial analyses between the 3D models corresponding to the different epochs become straightforward, as the targets materialize the same reference datum for all the 3D models.

At the same time, the installation, measurement and maintenance of a network of reference targets, firmly fixed on the seabed are time-consuming tasks, feasible only at limited depth and in warm and temperate waters. The peculiar conditions of underwater surveys in Antarctica with sub-zero temperatures and rigid safety procedures demand for a leaner method to allow the comparison of 3D photogrammetric models over time.

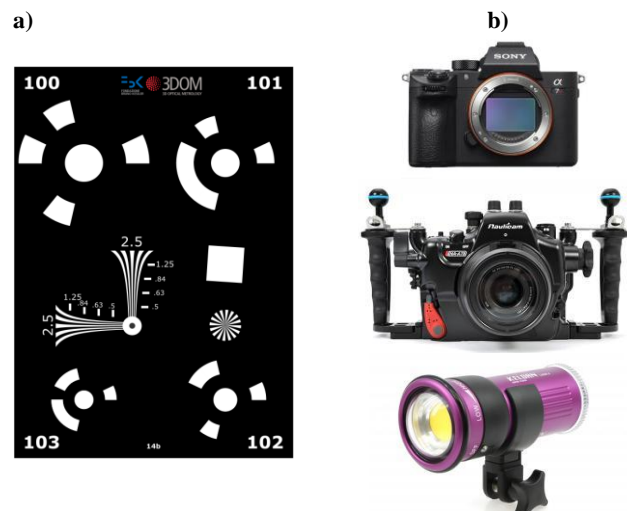
In previous works (Piazza et al., 2018; 2019 and 2020), the authors presented a first application of underwater photogrammetry in Antarctica to describe shallow-water rocky bottom benthic communities. 3D reconstructions of benthic habitats were obtained using both legacy video transects (i.e. not originally designed for photogrammetry) recorded in 2006 and 2015, and new videos from 2017, recorded using an updated protocol of image acquisition more suitable for photogrammetric measurements.

Originally, the videos from 2006 and 2015 were acquired respectively with a Sony HVR-HD1000E and HDR-HC7 HDV1080i interlaced camcorders in underwater housings mounting a flat port. These videos were just intended as a complementary tool (2006) or the main sampling tool (2015) for non-destructive visual surveys of benthic communities and habitats. Despite the limitations given by an older imaging technology based on interlaced videos, the non-ideal camera network geometry (single nadir looking strip) and the presence of a flat port, the study demonstrated the flexibility of photogrammetric techniques, capable of providing 3D metric results useful for ecological analyses with limited manual intervention.

In 2017, we updated the protocol of image acquisition using a stronger camera network geometry acquiring a central nadir and two oblique strips from each side (left and right) with 100% side lap. With this protocol, in collaboration with New Zealand and Korean SCUBA teams, new videos were recorded using a

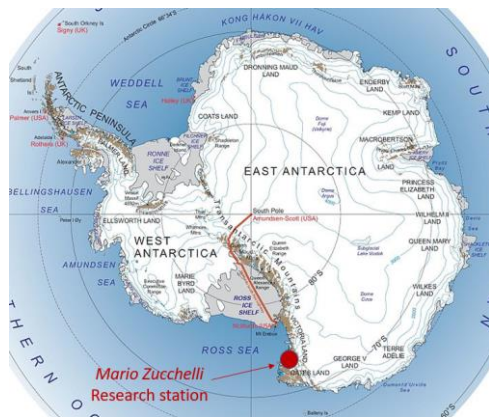
Sony A7sII mirrorless camera in a Nauticam Housing NA-A7II mounting a dome port. In depth benefits of this protocol are described in Piazza et al. (2018).

From 2018, to further automatize the scaling of the photogrammetric model, as well as to provide easier accuracy and reliability checks, the protocol included five target plates (Figure 1a) on the seabed, placed as temporary markers each time by the divers. The plates are made of printed Dibond in A5 format (148x210 mm) to make them durable and easy to carry underwater and show four circular coded targets and resolution patterns (wedges, square and siemens star). The resolution targets are used to check for the presence of motion blur as well as to verify that ground sample distance (GSD) is met during the survey. The multiple coded targets have a twofold purpose: (i) the known distances between the centres (accuracy better than 0.05 mm from calibration) are used as scale bars and (ii) the markers are triangulated after the bundle adjustment to verify that adjacent strips have been correctly oriented between each other. Although the surveying practices would suggest that scaling should be performed with lengths ideally comparable to the one to be surveyed, we estimate that, using multiple scalebars (more than eight with length of about 150 mm), a scaling accuracy better than 1:1000 (2 cm for 20m long transect) can be obtained in most situations.



**Figure 1.** Portable A5 target made in rigid Dibond aluminum with known distances and resolution targets used since 2018 for underwater photogrammetry campaigns in Antarctica (a). New camera system used from 2018 (Sony A7sII-16mm, Sony Vario-Tessar T\* FE 16-35mm F4) in a dedicated Nauticam housing with dome port mounting and led video light Keldan 4X Compact (10000 lumen) (b).

This paper reports new results from 2017, 2018, 2019 campaigns carried out with the new camera system (Figure 1, b) and the updated protocols of image acquisition specifically designed for photogrammetric purposes. In our previous papers (Piazza et al., 2018 and 2019) the single epochs were processed separately in a different coordinate system and only after co-registered to perform comparative analyses. Here we present a multi-temporal processing that provides co-registered 3D point clouds between the three years without control points nor direct georeferencing methods. This was possible thanks to the improved imaging technology and new camera network protocol as well as to the characteristics of the surveyed area where the presence of an extremely slow growing (estimated less than a mm per year) organisms (i.e., Corallinales, Rhodophyta algae), can be considered a



**Figure 2.** Italian research station “Mario Zucchelli” in Antarctica. Position on the map (left) and satellite view of its location along with the BTN\_PNRA\_T2 transect position (right).

natural stable pattern, useful for photogrammetric orientation of the different epochs. Moreover, for the first time we add a level of image understanding by introducing semantic segmentation in the process. Although semantic segmentation is ideally suited for thematic mapping and analyses on the photogrammetric products, such as for example on orthoimages (Pavoni et al., 2021) to estimate the coral reef growth, the underlying idea of using semantic segmentation in this paper is to further automatize and improve the quality control during the photogrammetric processing. Successively, the segmentation is also exploited into the generation of photogrammetric analyses and products (3D dense point clouds) to show how spatial analyses can be carried out on each class.

This represents an important test towards the development of fully automatic 3D analyses of benthic species in long-term ecological research programs.

In the next sections we briefly present the test areas and datasets, we then describe the processing steps undertaken to orient the single epochs in a single reference frame. A critical discussion and conclusion on the presented procedures will follow.

## 2. UNDERWATER PHOTOGRAMMETRY SURVEYS FOR LONG TERM ECOLOGICAL RESEARCH

### 2.1 A cooperative international team at “Mario Zucchelli” station

Underwater photogrammetry datasets acquired within the Italian national research program in Antarctica (PNRA) and used in this experimentation are collected within 2km from the Italian research base “Mario Zucchelli” (74°41'41.31”S 164°06'47.76”E), in the Terra Nova Bay area (Ross Sea - Figure 2). Thanks to the collaboration with New Zealand Researchers of Waikato University and the Korean Researchers of Korean

Polar Research Institute (KOPRI), scientific divers cooperate to maximize the underwater data collection during the different, often non overlapping periods of permanence at the research station. Divers enter the -1.8 °C polar water from a hole drilled in the ice about 3m thick in the month of November, when the water clarity is at its maximum and thus ideal for high resolution optical techniques like photogrammetry.

### 2.2 The chosen dataset and challenges in yearly transect revisiting

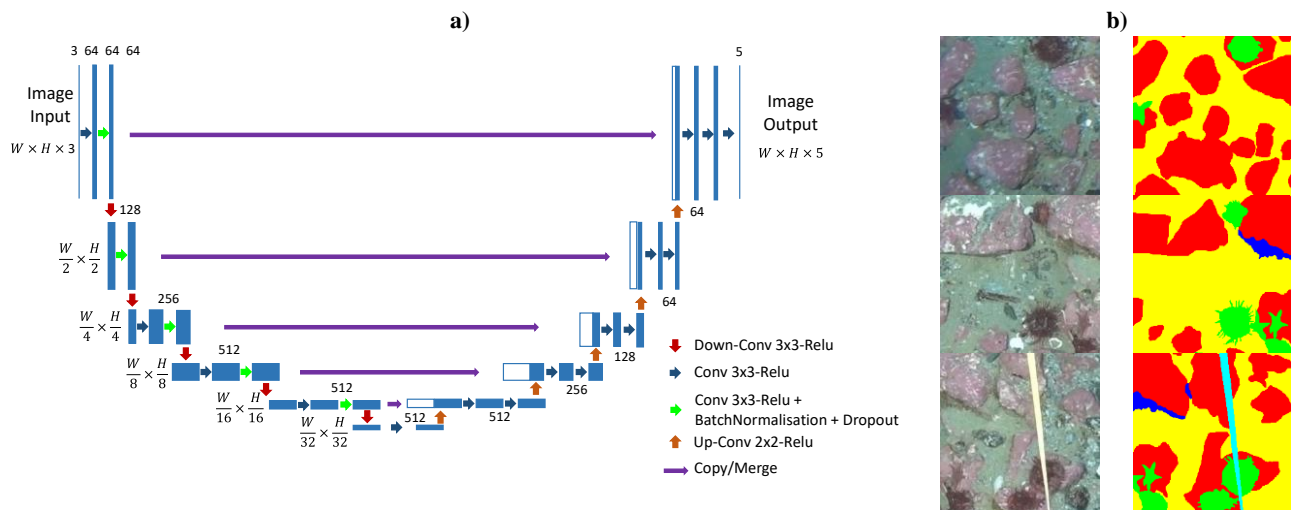
The dataset analysed in this paper (Table 1) is the BTN\_PNRA\_T2 named after the “Baia Terra Nova” (BTN). In particular we focus on transect T2 at a depth of about 20m. The transect does not have permanent targets materialized on the floor but only small concrete blocks and a iron block (segment of a train rail) that are used to help the diver recognize the extremities of the transect.

With a sub-millimeter ground sample distance (GSD) considered sufficient for the long term ecological research aims, Full HD (1920x1080) videos were recorded with the Sony A7sII with focal length set at 16mm at about 1m from the seabed. An average swimming speed between 15 and 25 cm/s was kept by the divers.

Videos are preferred by scientific divers for monitoring as they are easier to acquire and may be processed by following a variety of research protocols already in use. For the same reason, the autofocus system was left activated during the acquisition. The same team with same equipment carried out the photogrammetric acquisition in 2017, 2018, 2019. Unfortunately, only the 2018 dataset contains the target plates as in 2017 the protocol adopted did not consider their deployment yet.

Dataset Name	Date of acquisition	Target plates	GSD [mm]	Date Time Acquisition/ video duration		
				Left Strip	Nadir Strip	Right Strip
BTN_PNRA_2017_T2	15 November 2017	NO	0.9	2017:11:15 20:32:49 2m 20s	2017:11:15 19:24:18 1m 25s	2017:11:15 19:26:39 1m 26s
BTN_PNRA_2018_T2	12 November 2018	YES	0.8	2018:11:12 12:22:11 1m 36s	2018:11:12 12:19:26 1m 48s	2018:11:12 12:24:49 1m 47s
BTN_PNRA_2019_T2	16 November 2019	NO	0.8	not available	2019:11:16 12:03:47 1m 51s	not available

**Table 1.** Underwater photogrammetry dataset used for the experimentation.



**Figure 3.** Architecture of the U-NET convolutional neural network (CNN) used for semantic segmentation of the sea bottom (a). Example of training images (b) showing the RGB image patches and their corresponding ground truth (5 classes – see description in the text).

Moreover, budget constraints and limited diving personnel available for the different experiments in Antarctica may occasionally require that only the minimum necessary data be acquired. This was the case in 2019 when, for logistic and time constraints, target plates could not be deployed at the sea bottom and only the nadir acquisition was recorded.

This difference between the various datasets is one of the main characteristic and challenges in many long-term monitoring projects where teams and equipment may change over the years. As it is visible from the video metadata reported in Table 1, the three strips were recorded one after the other within few minutes with the exception of 2017, when the left strip was recorded after about one hour from the nadir looking and right strips.

These characteristics make the above videos a valuable dataset to develop photogrammetric methods and 3D data analyses that can work properly even in presence of unstable camera calibration (autofocus), improper camera model (flat or non-centered dome ports) and changes due to the dynamic nature of the scene.

### 3. METHODS

#### 3.1 Our approach to 4D monitoring of benthos leveraging underwater photogrammetric methods and image understanding

The benthos is a highly dynamic biotic component, that weakens the assumption of rigid body necessary in a standard photogrammetric process. In a multistrip survey the areas of overlap are revisited after a period of time during which changes at the seabed can happen. In extreme cases, changes can happen between successive images along the same strip. Indeed, depending on the specific habitat, the seabed might be almost completely covered by soft algae and marine phanerogams swinging in the current or crawling sea urchins and sea stars. Fish moving along and often following the photographer introduce other challenges and noise.

Using a deep learning approach based on the U-NET convolutional neural network (CNN), we trained and automatically segmented 5 classes as follows:

- sand, gravel and markers (yellow),

- pink coral alga *Tethysphytum antarcticum* Sciuto, Moschin & Moro 2021 (Sciuto et al., 2021), whose growth rate is estimated less than a mm/year (red),
- tape meter (cyan),
- crawling benthic species like sea urchins and sea stars (green),
- soft algae and animal turf (blue).

We use the semantically segmented classes with a twofold purpose within the photogrammetric workflow: (i) to guide the structure from motion (SfM) image orientation process by constraining the feature extraction on those areas of the images that are considered stable (i.e., excluding those species that move within a single survey, such as urchin and starfish); (ii) to analyse whether oblique adjacent strips have been correctly oriented with respect to the nadir one or that a specific epoch is well oriented with respect to the reference one.

#### 3.2 U-NET architecture

The proposed architecture for the segmentation is based on a U-NET network (Ronneberger et al., 2015).

The U-NET was proposed first by (Ronneberger et al., 2015) for fast and precise biomedical image segmentation. The architecture of a U-NET is composed of 2 symmetric paths: an encoding/contracting path to capture context and a decoding/expanding path that enables precise localization.

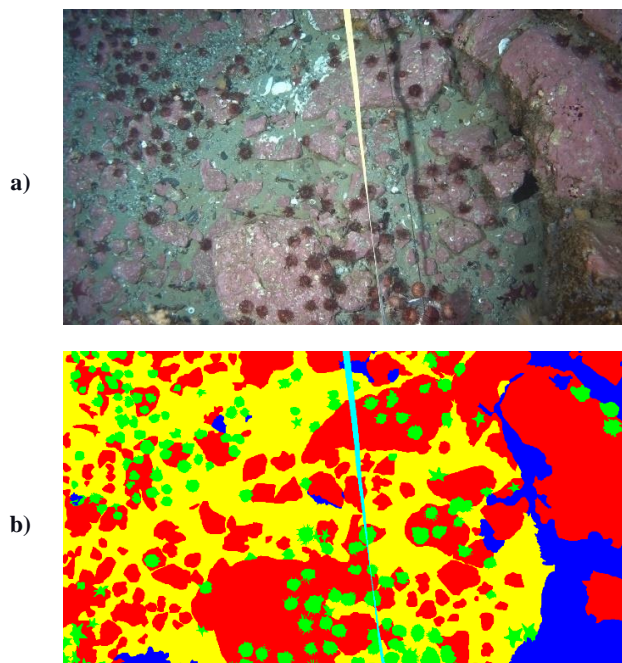
Figure 3a presents the proposed U-NET for the under-water image segmentation.

In our case, contracting part (left side of the network) has almost the same architecture of the VGG16 Convolutional Neural Network (Zisserman et al., 2015) with a series of 3x3 convolutions, followed by a rectified linear unit (ReLU), a batch-normalisation, a dropout and a 2x2 max pooling operation for the downsampling. Number of filters is equal to 64 at the first layer and is doubled at each of the next 3 downsampling steps. After that, it is maintained at 512 as maximum number of filters used in our network. In the expanding part (right side), feature maps are upsampled, followed by a 2x2 convolution that halves the number of feature channels (except for the first upsampling where the number of channels is maintained to 512). Unlike the fully convolutional approach, feature maps after each upsampling are concatenated with their

corresponding feature maps from the contracting path (Skip connexions), followed by further 3x3 convolutions and ReLu.

In our experiments, we labelled just one image. We divided it horizontally in 2 equal parts: the top part is used to create the training set and the other part (the bottom half) is used for testing. The training dataset is composed of 800 images of size  $256 \times 256$  cropped randomly from the half-top of the selected big image (Figure 3b).

In order to evaluate our method, we applied our trained model on the testing part of the image (Figure 4). The obtained accuracies over the area shown in Figure 4 are displayed in Table 2.



**Figure 4.** The original image (a) chosen for training and results of the testing used to evaluate the method (b). We divided it horizontally in 2 equal parts: the top part is used to create the training set and the other part (the bottom half) is used for testing.

Global Accuracy	97.51%	Global Accuracy	97.51%
Precision	97.61%	Precision	97.61%
F1-score	97.53%	F1-score	97.53%
IoU-score	94.46%	IoU-score	94.46%

**Table 2.** Statistics of the semantic segmentation on the test set.

### 3.3 Video frame extraction and SfM

The recorded videos were processed following the procedure described in Piazza et al. (2019). With the new photographic system characterized by a better sensor and more powerful led lights, the number of motion blurred frames significantly lowered. We simplified the procedure of frame extraction by

guaranteeing at least the 75% of overlap along the strip. This corresponded to a timed frame extraction at a frame rate of 2fps. The frames were then processed in Agisoft Metashape (www.agisoft.com/) where we added some additional functionalities to automatise the quality control (single strip dense cloud generation, scaling, across strip and multi epoch image orientation checks) via Python scripting. The different epochs are oriented together thanks to the particular characteristic of the recently described pink coral alga *Tethysphytum antarcticum* (Sciuto et al., 2021) whose growth rate is estimated less than a mm/year and thus represents a natural stable pattern that supports the automatic orientation between the different years.

For each epoch a different camera calibration was computed and available coded targets only marked after the bundle adjustment to evaluate the triangulation residuals on well signalled high contrast points. Inner-target known distances were then applied for scaling.

### 3.4 SfM in a dynamic environment with and without semantically masked images

We experimented the use of semantic segmentation to selectively mask the area of the images that are considered potentially unstable during the survey. Two scenarios were considered: (i) we suspect the scene to change within the same photogrammetric survey, thus only sand, gravel, markers (yellow class) and pink coral alga (red class) are considered unchanged; (ii) we consider long term monitoring applications, thus only the pink coral alga (red class) is considered stable across the different years. In the first case we tested the method on the 2018 images alone, in the second case we processed the three datasets 2017, 2018, 2019 together.

The masks are used to limit the feature point extraction only on the areas of the images labelled by classes considered stable. The idea is that, by eliminating the extraction and matching of potentially moving objects, the resulting image orientation would improve because it is less influenced by outliers. Moreover, by excluding parts of the images where feature points should not be trusted, the process becomes faster (between 30 and 70% in our experiments) and thus more efficient.

Table 3 reports the results of the two different scenarios each in terms of relative differences between the two data processing carried out with and without masks by analysing the precisions of the interior orientation parameters (we report the 2018 although the relative trend is comparable for 2017 and 2019), RMS of image residuals on the markers and length measurement error (LME) computed on 16 known lengths of the scale bars (inner target distance on the target plates).

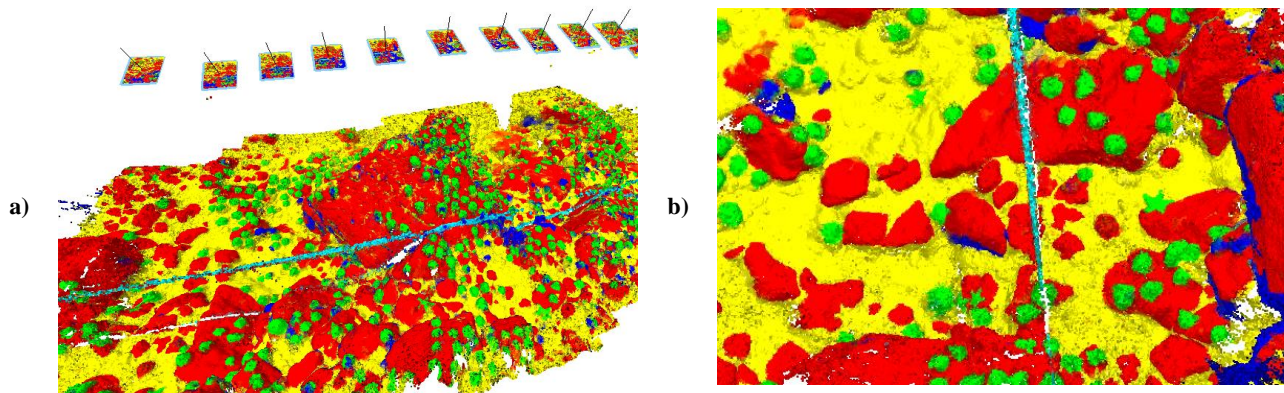
The results show that in the first scenario the masking slightly improves the results in object space (better LME).

On the contrary, a more significant worsening of precision and LME is observed for the scenario two when masks are used.

Most probably, if from one side potential outliers are removed from the processing when using the masks, on the other side there is a severe reduction in the number of image observations for scenario two, including good feature points wrongly excluded because of segmentation errors.

Processing version	$\sigma_c$ 2018 [px]	$\sigma_{PPAX}$ 2018 [px]	$\sigma_{PPAY}$ 2018 [px]	RMS target residuals [px]	LME [mm]
2018_masks_scenario_1	0.03	0.07	0.07	0.72	0.54
2018_no_masks_scenario_1	0.03	0.07	0.07	0.65	0.57
ALL_masks_scenario_2	0.04	0.08	0.09	0.70	1.01
ALL_no_masks_scenario_2	0.03	0.05	0.05	0.79	0.85

**Table 3.** Synthesis of the relative comparison between dataset processing with and without masking from semantic segmentation.



**Figure 4.** Multi-view semantic segmentation propagation from 2D to 3D. A 3D view of the segmented point cloud along with the camera poses in space (a) and a top view (b) of the point cloud as seen in the test image of Figure 4.

Overall, with respect to the previous papers (Piazza et al., 2018 and 2019) where legacy videos from 2006 and 2015 provided an LME of about 1.9 mm, we obtained an accuracy improvement up to three times.

### 3.5 Semantic segmentation to assess the quality of image orientation between adjacent strips

A successful image orientation would provide metrically co-registered 3D point clouds for the left, nadir and right strips when the dense matching is performed separately for each strip. In this case, the 3D point clouds should differ between them only by the expected triangulation accuracy. Strips not well connected with the others, due to the lack of tie points across the strips generate significant errors when the point clouds are compared with each other. These differences are even more evident in presence of systematic errors due to unstable camera calibration or improper camera model (Menna et al., 2020; Nocerino et al., 2021).

An automatic point-to-point comparison between the point clouds, separately generated for each adjacent strips, would also include differences due to actual changes caused by the dynamic characteristics of the scene.

With this in mind, the developed procedure for quality control automatically generates a different dense point cloud for each of the three strips that is semantically segmented by propagating the 2D semantic segmentation to the 3D point clouds using a multi-view approach (Riemenschneider et al., 2014; Stathopoulou and Remondino 2019). By leveraging the photogrammetric orientation of the images, the 3D points are back-projected on the images using the collinearity equations with a z-buffer visibility analysis using Matlab ([www.mathworks.com](http://www.mathworks.com)). The 2D image coordinates are then used to associate the corresponding class to the 3D point. As each 3D point is visible on multiple images, a median filter is used to eliminate spurious values due to back-projection and/or classification errors, finally assigning a unique class to the point. The points belonging to the classes considered stable during the survey within that specific year (e.g. sand and corals) are compared to each other using a point-to-point distance analysis. The nadir strip is considered the reference and the other two oblique are compared. Whenever a percentage (>5%) of points surpasses a predefined accuracy threshold we raise a warning requiring the results to be visually inspected and, consequently, the processing revised (for example by adding manual tie points between the strips).

Similarly, keeping only the class related to the pink coral algae *T. antarcticum*, quality checks can be performed to verify that

the image orientation has been performed correctly across the years. One epoch is chosen as reference and the others are compared relatively. By verifying that different strips are accurately oriented with respect to each other, we make sure that 3D point-to-point spatial comparisons only highlight actual changes between the different epochs, a fundamental requirement for ecological spatial analyses.

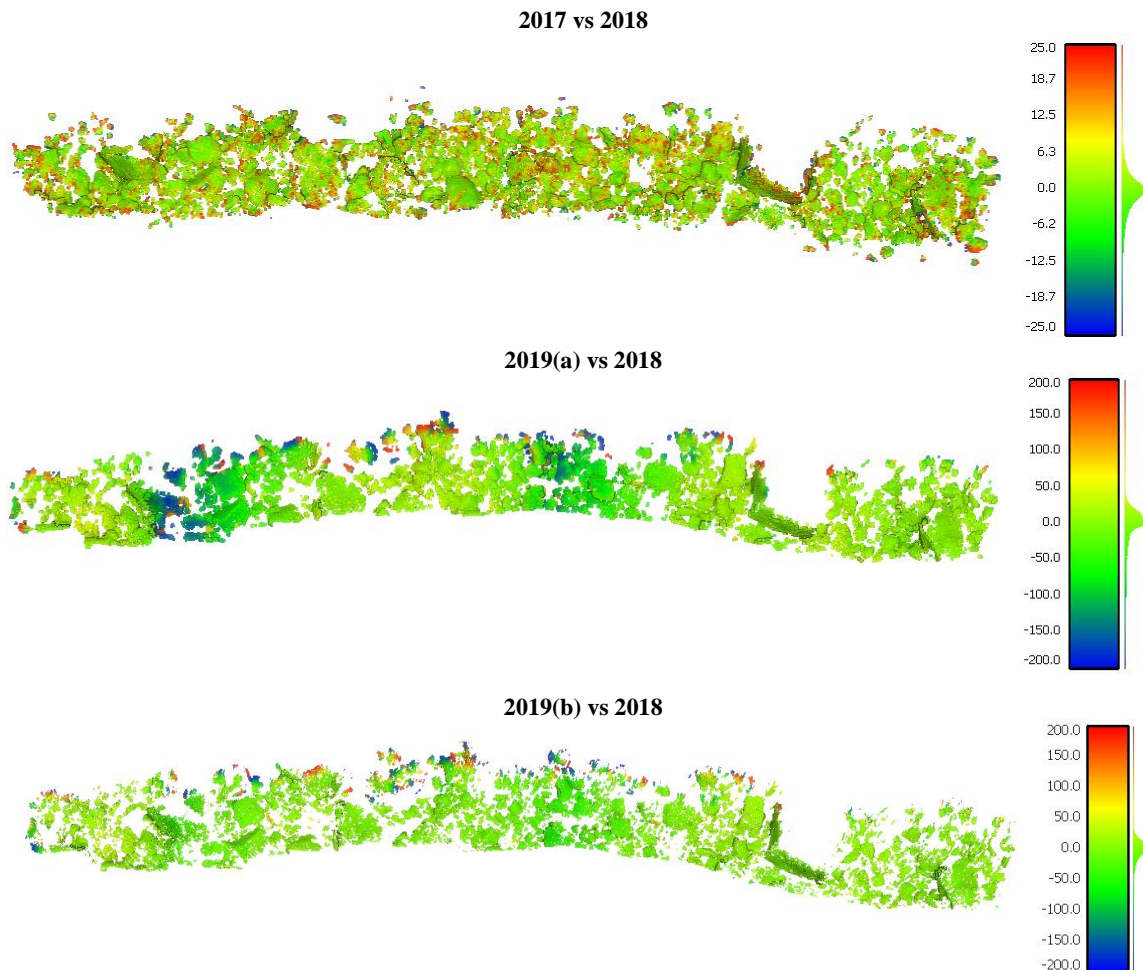
Figure 5 shows an example of application to the multi epoch quality control performed only on the rocks covered by the pink coral algae where the year 2018 is chosen as reference. The comparison between 2017 and 2018 shows no systematic effects and only a small percentage (<5%) of points differing by more than 3cm. On the other hand, the 2019 strip did not orient correctly showing some 3D polynomial deformations of the model up to 25 cm with 40% of the total number of points above the 3cm threshold (Figure 5, 2019a vs 2018).

After revising the processing and re-orienting the images from 2019 with four additional manually-marked and well-distributed points along the strip, the distances reduced similarly to what obtained with the 2017 epoch (i.e., only a small percentage <5% of points is more than 3cm from the reference) (Figure 5, 2019b vs 2018).

## 4. DISCUSSION

Quality checks are of utmost importance in surveying practices, having the delivered measurements a key role in any subsequent decision-making analyses and processes. For this reason, photogrammetric products require to be compliant with predefined project accuracy and reliability specifications, which are to be verified through a strict quality control procedure. In photogrammetry, accuracy and reliability checks are performed through inner measures and analyses of the bundle adjustment results (such as, for example the sigma naught, RMS of image observation residuals, precision of estimated parameters and their correlations, redundancy figures) and external analyses, i.e., from independent 2D/3D reference measurements in object space (control points, check points, scale bars). A suspect high RMS of image residuals from the bundle adjustment is a symptom that image observations do not properly fit the mathematical model.

This may be due to several reasons, such as wrongly measured and matched tie points (e.g., repetitive texture pattern, moving objects), inappropriate (e.g., rectilinear instead of fisheye lens) or an incomplete (e.g., flat port housing underwater, non-centered dome port, image stabilization, see Nocerino et al., 2022) calibration model for the camera.



**Figure 5.** Point-to-point comparison [mm] on the semantically segmented classes labelled as “stable” used as quality control across the different epochs.

Inner figures from the bundle adjustment do not always guarantee the required accuracy in object space, especially under a weak camera network geometry (single strip corridor mapping or the normal case in aerial photogrammetry). A warping in object space can be present even with subpixel RMS of image residuals (Menna et al., 2020) and can be measured using 3D check points, of known accuracy, distributed on the entire area. Unfortunately, control points are arduous to collect underwater and, in the practice of field operations, prohibitive in Antarctica.

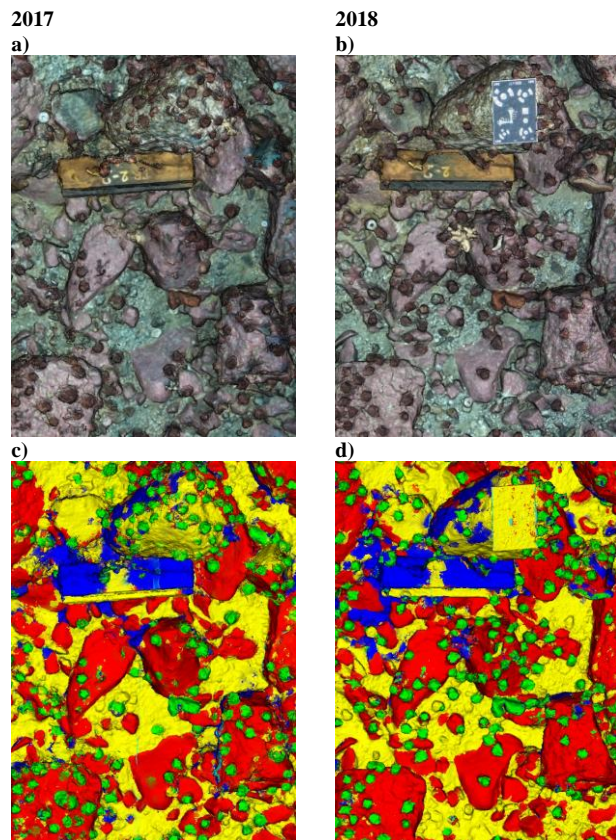
In depth analyses of photogrammetric results can be a hard task, especially in SfM where the high number of observations make it difficult to exploit all the statistical techniques used in traditional geodetic surveying. Moreover, SfM approaches include outlier removal strategies that may not be transparent to the user, often working as a black box. These procedures may remove an excessive number of image observations if their residuals are larger than a predefined threshold. The result is that even in presence of uncompensated systematic errors, such as due to a wrong camera calibration model (very frequent underwater) the RMS of image residuals is very low after the bundle adjustment, as only the observations that fit well the mathematical model within the given threshold are kept. The non-photogrammetrist may draw the wrong conclusion that image data have been correctly processed providing metrically correct results. For the 2019 strip the SfM process in Metashape ended with all the images oriented with subpixel residuals

together with the 2017-2018 images. Nevertheless, the orientation of 2019 images was wrong as shown in section 3.5, pointed out by the implemented quality control method. The issue can be ascribed to not enough tie points matched between the 2019 and the other two years. This corresponds to a weaker network geometry, consisting of a single nadir looking elongated strip and causing worse self-calibration parameters compared to 2017 and 2018 (higher standard deviations for the interior orientation and distortion parameters by more than an order of magnitude, e.g.:  $\sigma_c$  2018 = 0.03 pixels vs  $\sigma_c$  2019 = 0.3 pixels).

## 5. CONCLUSIONS AND FUTURE WORKS

In this paper we provided an overview of the latest updates on the photogrammetric procedures applied in Antarctica within the Italian long-term ecological research project for monitoring benthic species. The new underwater camera system with better sensor, optics and illumination provided video of superior image quality that allowed to improve the accuracy three times with respect to the previous legacy videos analyzed in Piazza et al., (2018, 2019). The new protocol of image acquisition with three strips (nadir looking and two obliques from the sides) and deployable target plates confirmed to significantly automatize the processing providing at the same time additional control in terms of length measurement error (LME) and check on the orientation between adjacent strips.

Also, we proved that, due to the presence of a coral algae *T. antarcticum* stable over the years, the different epochs can be oriented together even without the need of reference fixed markers on the sea bed allowing a direct comparison (Figure 6).



**Figure 6.** Example of 3D textured mesh models colored according to the RGB (a, b) and semantic segmentation (c, d) depicting the same areas from 2017 and 2018 obtained using the 2018 as reference epoch. Beside the evident changes in sea stars and urchin number and distribution, it can be noticed that the iron block slightly moved between the two epochs.

We experimented the use of semantic segmentation as additional step to provide image understanding within the data processing. This step has the twofold purpose of speeding up the image orientation and reducing potential outliers among the tie points by excluding the areas of the image belonging to parts of the benthos that can move during the survey, thus invalidating the rigid body assumption made in photogrammetry. The preliminary tests showed that current RANSAC techniques used in modern SfM methods are robust to the presence of a large number of outliers thus not evidencing a significant improvement in terms of accuracy when applying the semantic segmentation masking. We will further experiment the method on other datasets where the habitat is even more dynamic. Furthermore, we plan to use a stochastic model to weight the observations differently, according to their potential motion. The method will be implemented following the procedure already used in the past by the authors in Menna et al (2018).

We used the semantically segmented point clouds of the coral algae to verify the correct photogrammetric orientation of the different strips and epochs by comparing the dense point clouds belonging to adjacent strips of the same year or between different strips across the different years.

We will further refine the training of the U-NET network to include the different classes of interest (sea stars and urchins separated) for the ecological research also experimenting with instance segmentation for counting and measuring the single individuals.

The issues in applying the full protocol for 2019 campaign proved again the weakness of a single strip acquisition, which did not orient correctly showing severe deformation of the 3D point cloud that were resolved through manual tie points with the images belonging to the strips acquired in the other years.

To overcome the logistic difficulties, reduced bottom time when diving and limited scientific diving personnel available in Antarctica, we are planning to further improve the image acquisition protocol. To this scope, a new system is currently under development within the Italian National Antarctic Research Program (PNRA) research project “RosS-BMP” (Ross Sea Benthic Monitoring Program: new non-destructive and machine-learning approaches for the analysis of benthos patterns”). The system, consisting of a synchronized stereo camera featuring pressure and inertial sensor (Menna et al., 2021) will enable a higher redundancy in a single round trip passage over the transect instead of the three strips currently implemented. This system will thus reduce the bottom time underwater making the survey more efficient and safer. Moreover the presence of pressure and inertial sensor will allow a redundant system for scaling and levelling with respect to the local vertical direction.

#### ACKNOWLEDGEMENTS

This study has been partially supported by the project “Ross Sea Benthic Monitoring Program: new non-destructive and machine-learning approaches for the analysis of benthos patterns” (“RosS-BMP”,\_PNRA18\_00263). The authors are grateful to the New Zealand and Korean teams for their collaboration and precious contribution in the survey data acquisition. This paper is an Italian contribution to the CCAMLR CONSERVATION MEASURE 91-05 (2016) for the Ross Sea region Marine Protected Area, specifically, addressing the priorities of Annex 91-05/C.

#### REFERENCES

- Callahan, J.T., 1984. Long-term ecological research. *BioScience*, 34(6), 363-367.
- Chan, K.L., Qin K., 2017. Biomass burning related pollution and their contributions to the local air quality in Hong Kong. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2/W7, 29-36.
- Chennu, A., Färber, P., De’ath, G., de Beer, D. and Fabricius, K.E., 2017. A diver-operated hyperspectral imaging and topographic surveying system for automated mapping of benthic habitats. *Scientific reports*, 7(1), 1-12.
- Conlan, K.E., Kim, S.L., Lenihan, H.S. and Oliver, J.S., 2004. Benthic changes during 10 years of organic enrichment by McMurdo Station, Antarctica. *Marine Pollution Bulletin*, 49(1-2), 43-60.
- Deregibus, D., Quartino, M.L., Zacher, K., Campana, G.L. and Barnes, D.K., 2017. Understanding the link between sea ice, ice scour and Antarctic benthic biodiversity—the need for cross-station and international collaboration. *Polar Record*, 53(2), 143-152.



- Hosie, G., Mormede, S., Kitchener, J., Takahashi, K., Raymond, B., 2014. Near-surface zooplankton communities. In: Broyer C, Koubbi P (eds) *Biogeographic Atlas of the Southern Ocean. Scientific Committee on Antarctic Research*. Cambridge, 422–430.
- Menna, F., Nocerino, E., Chemisky, B., Remondino, F. and Drap, P., 2021. Accurate Scaling and Levelling in Underwater Photogrammetry with a Pressure Sensor. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLIII-B2-2021, 667–672.
- Menna, F., Nocerino, E., Ural, S. and Gruen, A., 2020. Mitigating image residuals systematic patterns in underwater photogrammetry. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLIII-B2-2020, 977–984
- Menna, F., Nocerino, E., Drap, P., Remondino, F., Murtiyoso, A., Grussenmeyer, P. and Börlin, N., 2018. Improving underwater accuracy by empirical weighting of image observations. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, pp.699–705.
- Nocerino, E., Menna, F. and Verhoeven, G., 2022. Good vibrations? How image stabilisation influences photogrammetry. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 46, pp.395–400.
- Nocerino, E., Menna, F., Gruen, A., Troyer, M., Capra, A., Castagnetti, C., Rossi, P., Brooks, A.J., Schmitt, R.J. and Holbrook, S.J., 2020. Coral Reef Monitoring by Scuba Divers Using Underwater Photogrammetry and Geodetic Surveying. *Remote Sensing*, 12(18), p.3036.
- Parravicini, V., Morri, C., Ciribilli, G., Montefalcone, M., Albertelli, G. and Bianchi, C.N., 2009. Size matters more than method: visual quadrats vs photography in measuring human impact on Mediterranean rocky reef communities. *Estuarine, Coastal and Shelf Science*, 81(3). 359–367.
- Pavoni, G., Corsini, M., Ponchio, F., Muntoni, A., Edwards, C., Pedersen, N., Sandin, S. and Cignoni, P., 2021: TagLab: AI-assisted annotation for the fast and accurate semantic segmentation of coral reef orthoimages. *Journal of Field Robotics*.
- Peirano, A., Bordone, A., Marini, S., Piazza, P. and Schiaparelli, S., 2016. A simple time-lapse apparatus for monitoring macrozoobenthos activity in Antarctica. *Antarctic Science*, 28(6), 473–474.
- Piazza, P., Gattone, S.A., Guzzi, A., Schiaparelli, S., 2020. Towards a robust baseline for long-term monitoring of Antarctic coastal benthos, *Hydrobiologia*, 847: 1753–1771.
- Piazza P., Cummings V., Guzzi A., Hawes I., Lohrer D., Marini S., Marriott P., Menna F., Nocerino E., Peirano A., Kim S., Schiaparelli S. 2019. Underwater photogrammetry in Antarctica: long-term observations in benthic ecosystems and legacy data rescue. *Polar Biology*, 42: 1061–1079.
- Piazza, P., Cummings, V.J., Lohrer, D.M., Marini, S., Marriott, P., Menna, F., Nocerino, E., Peirano, A. and Schiaparelli, S., 2018. Divers-operated underwater photogrammetry: Applications in the study of antarctic benthos. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, pp.885–892.
- Pizarro, O., Friedman, A., Bryson, M., Williams, S.B. and Madin, J., 2017. A simple, fast, and repeatable survey method for underwater visual 3D benthic mapping and monitoring. *Ecology and Evolution*. 7(6).1770–1782.
- Riemenschneider, H., Bódis-Szomorú, A., Weissenberg, J. and Gool, L.V., 2014. Learning where to classify in multi-view semantic segmentation. In European Conference on Computer Vision (pp. 516–532). Springer, Cham.
- Ronneberger, O., Fischer, P. and Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention, pp. 234–241. Springer, Cham.
- Sciuto, K., Moschin, E., Alongi, G., Cecchetto, M., Schiaparelli, S., Caragnano, A., Rindi, F. and Moro, I., 2021. *Tethysphytum antarcticum* gen. et sp. nov.(Hapalidiales, Rhodophyta), a new non-geniculate coralline alga from Terra Nova Bay (Ross Sea, Antarctica): morpho-anatomical characterization and molecular phylogeny. *European Journal of Phycology*, pp.1–12.
- Simonyan, K. and Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
- Smith, R.C., Baker, K.S., Fraser, W.R., Hofmann, E.E., Karl, D.M., Klinck, J.M., Quetin, L.B., Prézelin, B.B., Ross, R.M., Trivelpiece, W.Z. and Vernet, M., 1995: The Palmer LTER: A long-term ecological research program at Palmer Station, Antarctica. *Oceanography*, 8(3), 77–86.
- Solan, M., Germano, J.D., Rhoads, D.C., Smith, C., Michaud, E., Parry, D., Wenzhöfer, F., Kennedy, B., Henriques, C., Battle, E. and Carey, D., 2003. Towards a greater understanding of pattern, scale and process in marine benthic systems: a picture is worth a thousand worms. *Journal of Experimental Marine Biology and Ecology*, 285. 313–338.
- Stark, J.S., Snape, I. and Riddle, M.J., 2006. Abandoned Antarctic waste disposal sites: monitoring remediation outcomes and limitations at Casey Station. *Ecological Management & Restoration*, 7(1), 21–31.
- Stathopoulou, E.-K. and Remondino, F., 2019. Multi-view stereo with semantic priors. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2/W15, 1135–1140.
- Strayer, D., Glitzenstein, J.S., Jones, C.G., Kolasa, J., Likens, G.E., McDonnell, M.J., Parker, G.G. and Pickett, S.T., 1986. Long-term ecological studies: an illustrated account of their design, operation, and importance to ecology. *Occasional publication of the Institute of Ecosystem Studies*, 2(9).
- Trygonis, V. and Sini, M., 2012. photoQuad: a dedicated seabed image processing software, and a comparative error analysis of four photoquadrat methods. *Journal of Experimental Marine Biology and Ecology*, 424, pp.99–108.