

A Trillion Coral Reef Colors: Deeply Annotated Underwater Hyperspectral Images for Automated Classification and Habitat Mapping

Ahmad Rafiuddin Rashid and Arjun Chennu *

Max Planck Institute for Marine Microbiology, Celsiusstrasse 1, 28359 Bremen, Germany;
raf.rashid@outlook.com

* Correspondence: achennu@mpi-bremen.de

Received: 28 January 2020; Accepted: 17 February 2020; Published: 18 February 2020

Abstract: This paper describes a large dataset of underwater hyperspectral imagery that can be used by researchers in the domains of computer vision, machine learning, remote sensing, and coral reef ecology. We present the details of underwater data acquisition, processing and curation to create this large dataset of coral reef imagery annotated for habitat mapping. A diver-operated hyperspectral imaging system (HyperDiver) was used to survey 147 transects at 8 coral reef sites around the Caribbean island of Curaçao. The underwater proximal sensing approach produced fine-scale images of the seafloor, with more than 2.2 billion points of detailed optical spectra. Of these, more than 10 million data points have been annotated for habitat descriptors or taxonomic identity with a total of 47 class labels up to genus- and species-levels. In addition to HyperDiver survey data, we also include images and annotations from traditional (color photo) quadrat surveys conducted along 23 of the 147 transects, which enables comparative reef description between two types of reef survey methods. This dataset promises benefits for efforts in classification algorithms, hyperspectral image segmentation and automated habitat mapping.

Dataset: <https://doi.org/10.1594/PANGAEA.911300>

Dataset License: CC-BY-NC

Keywords: hyperspectral imaging; proximal sensing; machine learning; hierarchical learning; coral reef; biodiversity; classification; habitat mapping; image segmentation

1. Summary

Assessing coral reef habitats has historically been difficult because they are highly heterogeneous and structurally complex systems. Reef habitat structures can vary substantially, both between and within reefs, in terms of topography, water depth, community composition, and remoteness from human populations [1]. Comprehensive assessments of large areas of reefs are required to sufficiently sample and represent the level of this heterogeneity. While in situ benthic field survey methods have long been the gold standard among coral reef ecologists and widely deployed for many monitoring programs around the world, they are limited in spatial scale due to logistical constraints [2,3]. Furthermore, taxonomic identification by experts is time-consuming, expensive, biased [4], and can delay the availability of survey data where near real-time monitoring data is preferred [5]. Therefore, there is a pressing need for a rapid and scalable method of assessing coral reefs. It is within this context that close-range [6], or underwater hyperspectral imaging [7] has been developed and deployed for surveying the habitat structure of coral reefs.

Hyperspectral imaging has proven to be a powerful remote sensing technique, with many different applications in agriculture [8], forestry [9], urban planning [10], and ecology [11].

Hyperspectral imaging refers to the collection of optical images across a wide range of wavelengths in hundreds of narrow contiguous bands [12], as opposed to 3 widely separated channels in color photography. Remote sensing analyses have elaborated automated classification tools to exploit the high spectral resolution of hyperspectral images and map large areas of land and ocean. Coral reef mapping using hyperspectral imaging has mostly employed airborne and satellite platforms [13–16], resulting in maps of limited spatial and taxonomic resolution. Proximal sensing with an underwater system brings the promise of increased spectral signal fidelity and higher spatial resolution, enabling better biodiversity identification and classification to much higher taxonomic resolutions [7].

The dataset described in this paper contains benthic survey data from coral reefs around the Caribbean island of Curaçao. Co-located underwater hyperspectral imagery and color photographs of 23 different scenes across 8 coral reef sites at different depths have been annotated for habitat descriptors and benthic taxonomic identity, down to mostly genus- and species-levels. We formulated a novel protocol to annotate hyperspectral coral reef imagery to reduce the amount of manual identification, thus easing the data annotation effort. This dataset also contains an independent set of annotations to develop and validate image segmentation efforts to extract semantic descriptions of habitat maps. Beyond this, a large part of the dataset contains unannotated hyperspectral imagery for automated classification and habitat mapping.

Given the high taxonomic resolution of our annotations, this is perhaps the most detailed publicly available dataset for reef habitat mapping. Coral reef ecologists would benefit from the availability of this dataset, as it can be used to develop tools for scalable habitat description. This type of data is also of interest to research communities seeking real-world datasets to improve machine learning workflows for automated analyses such as data fusion, classification and segmentation [17]. As our annotation labels are hierarchically linked, hierarchical paradigms can be explored with this dataset. Another avenue of interest may be the use of incremental learning, where data is consumed gradually by classifier models, since this dataset contains hyperspectral images distributed across water depths (between 3 m and 10 m) and geographical location [18].

2. Data Description

The dataset comes from surveys, conducted between 4th and 26th of August 2016, of 147 underwater scenes of coral reefs around the Caribbean island of Curaçao. Each scene, approximately 50 m long and 1 m wide, was imaged underwater as a linear transect with a pushbroom hyperspectral imager in the HyperDiver system. Even though hyperspectral imaging of each 50 m transect took only a few minutes, transects across 8 sites were surveyed over different days and at different times of the day, thus covering a wide range of natural light settings. Across 147 hyperspectral images, there are a total of 2.29 billion pixels, each containing spectral data of 400 wavelengths, adding up to almost 1 trillion “colors” that reflect the scope of the dataset. Of these, 23 are comparative transects, because a series of color photographs was taken along the length of the transect. By annotating the co-located hyperspectral and color images from these transects, a comparative analysis could be performed. Separately, the hyperspectral images of 8 other transects were also annotated, to help independent assessment of benthic target segmentation (Section 2.3.1). Annotations were made by marking regions-of-interest (ROIs) as polygons in the 31 annotated hyperspectral images, and associating labels with these ROIs. These annotations contain 47 unique labels of habitat descriptors and biodiversity. Due to the plan-view nature of the survey methods, only sessile biota, substrate types and other abiotic items were identified. No fish census was carried out during the surveys.

2.1. Photo Quadrat Survey Data

The dataset contains transects of 50 m length where a plan-view color photograph (“photo quadrat”) was captured every 2 m along the transect. Of the 24 transects captured with this method, one was excluded due to data corruption of the corresponding hyperspectral image, resulting in 23 transects with a total of 575 images. In each photo quadrat, 80 random points were annotated for visible organisms or substrate found on the seafloor (Section 3.3.1)

2.2. Hyper Diver Survey Data

The HyperDiver is a diver-operated hyperspectral surveying system that can be used to capture high-resolution underwater transects of seafloor scenes with standard diver survey protocols [7]. A HyperDiver survey at each transect resulted in one hyperspectral image of the scene. Each image was rectified through visual estimation and the hyperspectral data were organized into a cube with three dimensions (y, x, w), where y and x represent the two spatial dimensions of the image, while w represents the spectral dimension. The spectra were interpolated to a standardized 400 channels between 400 nm and 800 nm. The HyperDiver also simultaneously measured downwelling irradiance of photosynthetically active radiation (PAR), depth and altitude profiles (Figure 1) as well as a high-resolution video stream of the scene. These supplementary sensor data give each HyperDiver transect some additional contextual information such as the variable light fields and topographic profiles.

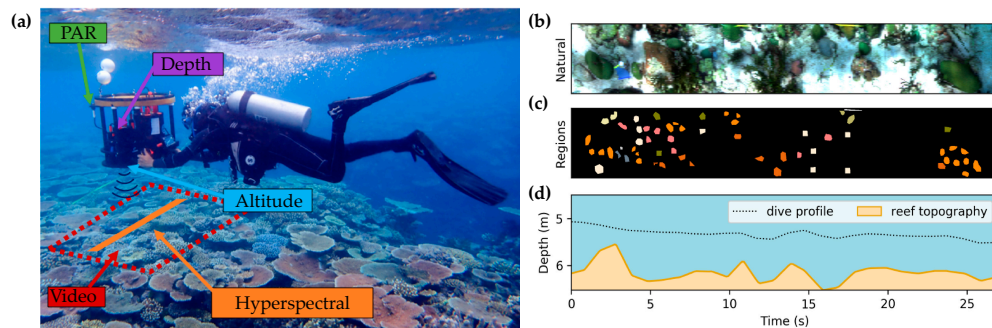


Figure 1. (a) The HyperDiver during an underwater survey of a coral reef transect. Sensors collect hyperspectral images, altitude, depth and irradiance. (b) Three-channel ‘natural’ view of the surveyed transect area derived from hyperspectral data. (c) Class map showing colored polygons that correspond to annotated regions-of-interest. (d) Depth and altitude information were used to generate the altitude and topographic profile of the transect.

2.3. Benthic Habitat Description

Among the 147 transects of HyperDiver survey data, we annotated a subset of them for benthic taxonomic identity, up to genus and species levels. The data from the two survey methods (HyperDiver and photo quadrats) were annotated separately, and with different annotation strategies. However, the hierarchical labels that had been used for the annotations were maintained across the two survey methods (See *Methods*). For the HyperDiver survey data, ROIs were manually selected and annotated with a class label, based on expert identification on a corresponding video. A total of 2089 ROIs were annotated across the 23 comparative transects, making up 8.2 million annotated data points of spectral information (Table 1). The average spectrum for a subset of the annotated classes were extracted for comparison (Figure 2).

Table 1. Curaçao coral reef sites surveyed using the HyperDiver and a tally of annotations by site.

Site Name	Site Location	Total Transects	Annotated Transects	Annotated ROIs	Annotated Pixels
Carmabi	12.122331°N, 68.969234°W	22	3	331	828,968
Kokomo	12.160331°N, 69.005403°W	20	3	244	968,617
Playa Kalki	12.375344°N, 69.158931°W	20	3	183	828,019
Habitat	12.197850°N, 69.079558°W	22	3	231	775,872
Water Factory	12.109989°N, 68.956258°W	10	3	377	1,347,646
Marie Pampoer	12.091894°N, 68.907918°W	18	3	281	1,076,596
Sea Aquarium	12.083234°N, 68.895114°W	15	2	117	1,117,412
East Point	12.042249°N, 68.745104°W	20	3	325	1,264,642
Total		147	23	2,089	8,207,772

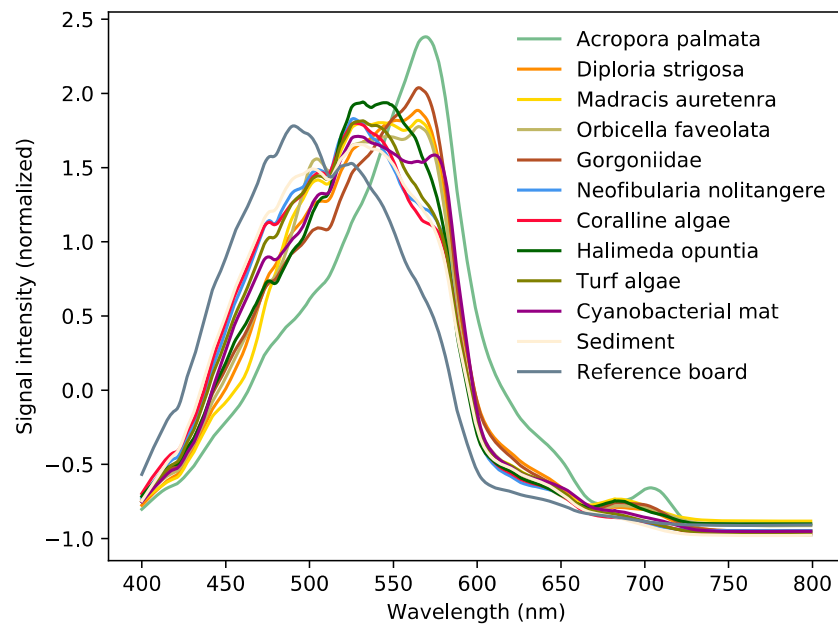


Figure 2. The normalized average spectrum shows the unique shapes and potentially distinguishing features of 12 different class labels. The average for each class label was calculated from all pixels annotated with that label across the dataset. These 12 labels provide a comparison between examples of labels occurring in the broader categories of corals, sponges, macroalgae and other habitat descriptors.

2.3.1. Targets for Image Segmentation

Apart from the annotated ROIs in the 23 comparative transects, another 56 ROIs were separately marked and annotated (Table 2). These annotations represent either whole colonies or contiguous areas of one particular target class. This set of ROIs can be useful to develop and assess image segmentation capabilities useful for habitat mapping. Image segmentation is a computer vision technique that subdivides an image into groups of pixels with a common semantic description (i.e. identity). In the context of reef habitat mapping, good image segmentation should result in machine generated semantic segments which correspond to habitat-level descriptions, such as full coral colonies, or contiguous areas of accurately labeled substrate.

The segmentation ROIs were annotated across 8 previously unlabeled transects (Transects 28, 46, 82, 90, 107, 125, 132 and 141) to provide an additional layer of validation by assessing how well a classifier is able to map unseen data. The target classes for these ROIs were chosen to represent a variety of significant categories or morphological types (Table 2). For example, in the Caribbean, branching corals of genus *Acropora* are considered indicators of reef health, due to their sensitivity to

environmental change [19,20]. Furthermore, the structural complexity of *Acropora* is likely to pose a different segmentation challenge than more evenly textured massive corals like *Siderastrea siderea*. The ability to segment full colonies of varying structural complexity (different morphologies, shapes, etc.) is an important metric to assess the quality of any reef mapping workflow. Rarely occurring class labels like *Acropora cervicornis* and *Aiolochoia crassa* have also been included among segmentation ROIs to help assess how class imbalances in the training data will affect the eventual performance of the automated habitat mapping. Across the different target class labels, with the exception of *Acropora*, there are at least three ROIs available for assessing segmentation and other analyses.

Table 2. Description of the category and target classes of the ROIs annotated for image segmentation.

Category	Sub-Category / Morphology	Target Class	Annotated Regions
Coral	Branching	<i>Acropora cervicornis</i>	2
		<i>Acropora palmata</i>	1
		<i>Madracis aurentenra</i>	3
Coral	Massive/sub-massive	<i>Diploria strigosa</i>	3
		(<i>Pseudodiploria strigosa</i>)	
		<i>Montastrea cavernosa</i>	3
		<i>Orbicella faveolata</i>	3
		<i>Orbicella annularis</i>	4
		<i>Siderastrea siderea</i>	3
Hydrozoan		<i>Millepora</i> sp.	3
Macroalgae	Brown	<i>Dictyota</i> sp.	4
Macroalgae	Green	<i>Halimeda opuntia</i>	3
Soft coral		Gorgoniidae	3
		Plexauridae	4
Sponge	Barrel	<i>Neofibularia nolitangere</i>	4
		<i>Ircinia campana</i>	4
Sponge	Massive	<i>Aiolochoia crassa</i>	3
Substrate		Sediment	3
		Coral rubble	3
Total			56

3. Methods

3.1. Data Acquisition

Survey data were obtained from 147 transects around Curaçao (12.166°N, −68.966°W) in the Caribbean Sea. In total, 8 coral reef sites all along the south-western coast of Curaçao were selected to cover a variety of habitat types with varying proximity to coastal urban settlements. Starting from the northernmost site, and in the southeast direction, the sites were Playa Kalki, Habitat, Kokomo, Carmabi, Water Factory, Marie Pampoen, Sea Aquarium, and East Point (Table 1; Figure 3).

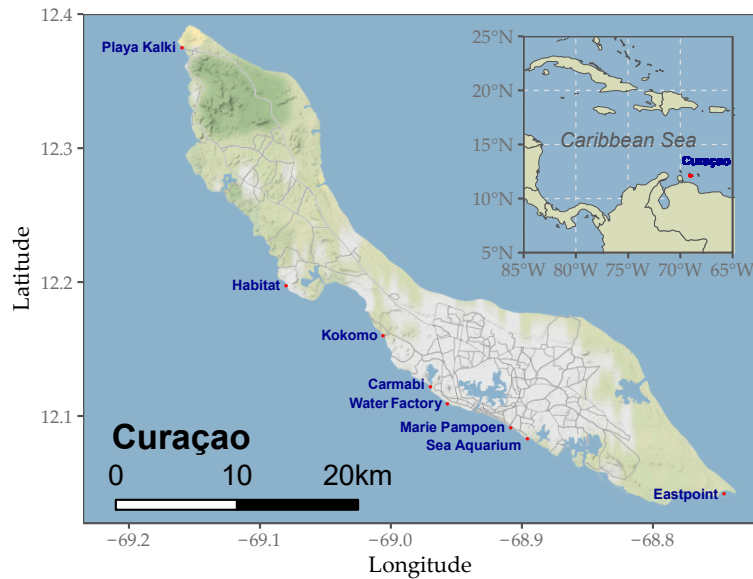


Figure 3. Locations of the 8 HyperDiver survey sites — Playa Kalki, Habitat, Kokomo, Carmabi, Water Factory, Marie Pampoën, Sea Aquarium, Eastpoint — around the island of Curaçao in the Caribbean Sea (inset).

At each of these sites, the HyperDiver was used to survey a large area by capturing between 10 and 20 hyperspectral transects in total (Figure 4). The surveyed area was marked by laying out 50 m measuring tapes along the edges of the area. The seafloor area between depths ~3 m to 9 m was surveyed in a raster pattern at mostly constant seafloor depth. Of these transects, only three were chosen as comparative transects, where co-located color imagery was also captured. These comparative transects were at three different depths—3 m, 6 m, and 9 m.

Every hyperspectral image of a transect includes a gray plastic board (25 cm × 25 cm) placed on the reef bottom at either one or both ends of the transect. This board provides a reference scale for spatial dimensions and the neutral gray color also allows estimation of the solar spectrum at a given depth and to derive local pseudo-reflectance values. The HyperDiver is maneuvered at near constant altitude along the direction parallel to the transect tape, ensuring that movement was largely smooth and compensated against cross-directional currents. During the scan, the HyperDiver continuously gathered a suite of different data at the same time (Figure 1; Section 3.2).

To acquire the co-located photo quadrat survey data, a diver captured plane-view color images with an underwater digital camera at constant altitude along the length of the transect at 2 m intervals (Figure 4). This results in the acquisition of a total of 25 images per comparative transect.

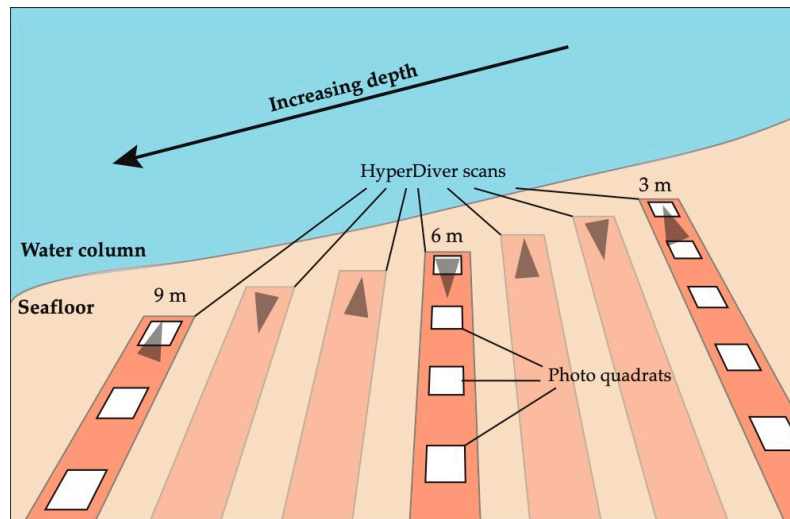


Figure 4. Each coral reef survey site was surveyed using two methods. Depending on the slope of the site, between 10 and 20 HyperDiver scans were carried out per site, in alternating directions (gray arrows). Three transects at different depths were also surveyed using traditional photo quadrat method, where one natural color image was captured at 2 m intervals along the transect (white). Only transects with co-located data were annotated for biodiversity and substrate type.

3.2. Hyper Diver Data Processing

Hyperspectral data was first manually inspected for data quality. Transects with corrupted or incomplete data ($N = 43$) were excluded. Transects deemed acceptable ($N = 147$) were rectified to represent the imaged scene as approximately square pixels. Rectification is necessary because the HyperDiver generates variable longitudinal and transverse resolution based on the imaging optics, swimming speed and frame acquisition rate [7]. Furthermore, since it is an underwater system without georeferencing capabilities, image rectification could not be easily automated. The rectification was performed through cropping unwanted sections at the ends of the captured transect scene as well as stretching of the scan in the y direction to produce nearly square pixels and a visually coherent image of the scene. The spectrum in each pixel was linearly interpolated to 400 bands in the 400 nm to 800 nm wavelength range, and its intensity scaled down from 16-bit to 8-bit radiometric resolution. The supplementary sensor data from the HyperDiver scan were also included in the dataset, but the use of these data should be approached with care as they have not been evaluated beyond calibration. These include photosynthetically active radiation (PAR), altitude, pressure, and system pose data like pitch, roll, yaw and acceleration. An underwater video camera was also mounted onto the HyperDiver to capture high-definition videos of the scene during each scan. As part of HyperDiver data processing, these videos were color-corrected to ease taxonomic identification. Video frames were extracted at 3 s intervals and are included in this dataset.

3.3. Biodiversity and Substrate Labels for Habitat Mapping

We created this dataset to be used for machine learning and automated habitat mapping. In order to produce a suitable learning dataset, we formulated a protocol to annotate a subset of the hyperspectral imagery with labels that are useful for habitat mapping, namely taxonomic identities for reef organisms and substrate descriptions for abiotic structures. We relied on existing ontologies, to propagate standard vocabularies of habitat description and to avoid creating a new annotation schema. For labels of reef organisms, we adopted a well-established database of marine organisms, the World Register of Marine Species (WoRMS) [21]. WoRMS provides scientific names of marine taxa, through unique and stable identifiers via its 'Aphia' database model [22]. For abiotic habitat substrate labels, we relied on the CATAMI classification scheme [23], in an effort to keep it consistent with best-practices in marine imagery annotations. Both the CATAMI and Aphia schemes of labeling

are hierarchical, and may enable the use of hierarchical learning paradigms for classification tasks (Figure 5). Combining these two annotation schemes, we were able to assign up to three tags per annotated data point, namely ‘visual’, ‘worms’ and ‘catami’. At the very least, an annotation consists of only a *visual* tag if it is neither a reef organism nor an abiotic substrate (e.g., “turf algae”). To label abiotic classes, both the *visual* and *catami* tags were assigned, like in the case of *visual* “sediment”, which was also tagged with a *catami* tag of “82001005” corresponding to soft, unconsolidated substrate under the CATAMI scheme. For biotic annotations, we assigned all three tags to it e.g. the *visual* “*Siderastrea siderea*” also has a *worms* tag of “207516” and *catami* tag of “11290906” for massive hard corals. In total, we generated 47 unique *visual* labels across the annotated dataset.

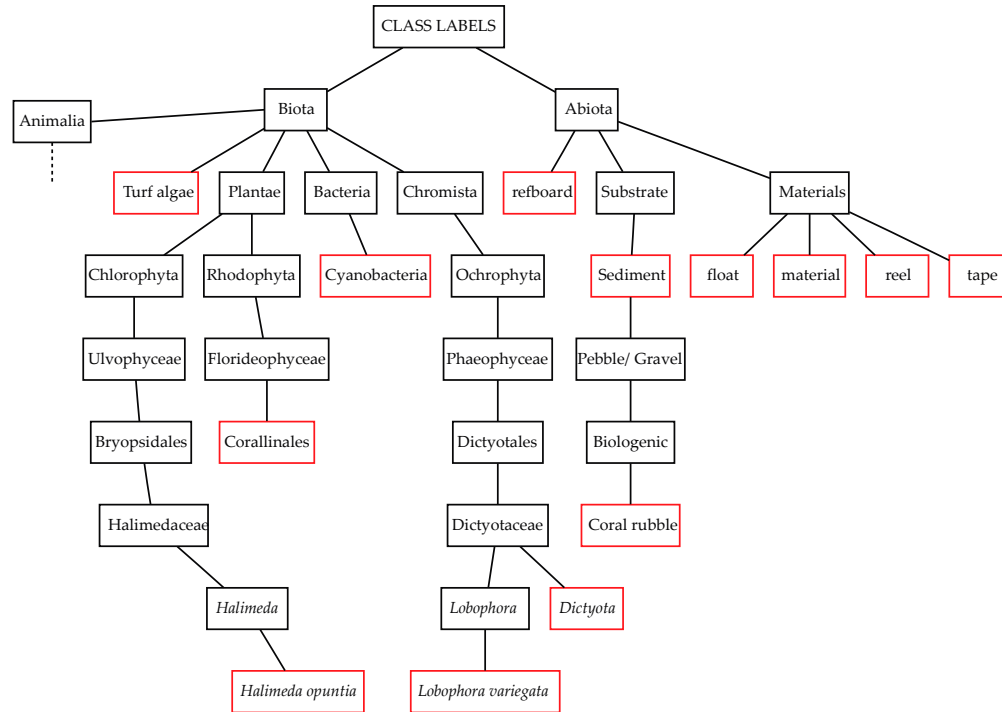


Figure 5. Truncated tree schematic showing hierarchical nature of annotation labels was reconstructed from the unique identifiers from existing classification ontologies. Labels used for the annotated dataset are in red bounding boxes. Biotic labels were tagged with unique AFAID identifier from the World Register of Marine Species (WoRMS) database and abiotic labels were tagged with an identifier code from the Collaborative and Annotation Tools for Analysis of Marine Imagery and video (CATAMI) classification scheme. For the complete hierarchical tree that includes all labels under Kingdom Animalia, refer to Figure A1 in Appendix A.

3.3.1. Annotation Strategy: Random Point Count Method

Determining coral and substrate coverage in a photo quadrat employs the random point count methodology, a widely used method of obtaining minimally-biased estimations of percent cover from image-based surveys [24,25]. Using a point annotation software called Coral Point Count with Excel extensions (CPCe) [25], 80 random points were generated on each of the 25 photo quadrats. Every randomly selected point was identified either as a substrate type or a benthic organism to the highest taxonomic resolution possible, resulting in a total of 2000 annotated data points per transect. Pixel coordinates of the annotated points were then saved with their respective labels as a table (.csv file) associated with each quadrat. Since expert annotations are time-consuming and expensive to obtain, the reduced effort is achieved through labeling only a small subset of random points on each image [19]. For this dataset, 80 random points were chosen based on previous research [19] which used

simulated data to determine that for a reef transect with 20% hard coral cover, 80 points per image was the minimum number of points to obtain for accurate percent cover estimates.

3.3.2. Annotation Strategy: Deliberate Bias to Reduce Human Effort

For the annotation of HyperDiver data, a natural color rendering of each transect's hyperspectral image was visualized and annotated interactively with ROI polygons, with the help of the accompanying video recording of the scene (Figure 6). As opposed to the previously described random point selection method for photo quadrat surveys (Section 3.1), we developed an annotation strategy for HyperDiver data that was deliberately biased.

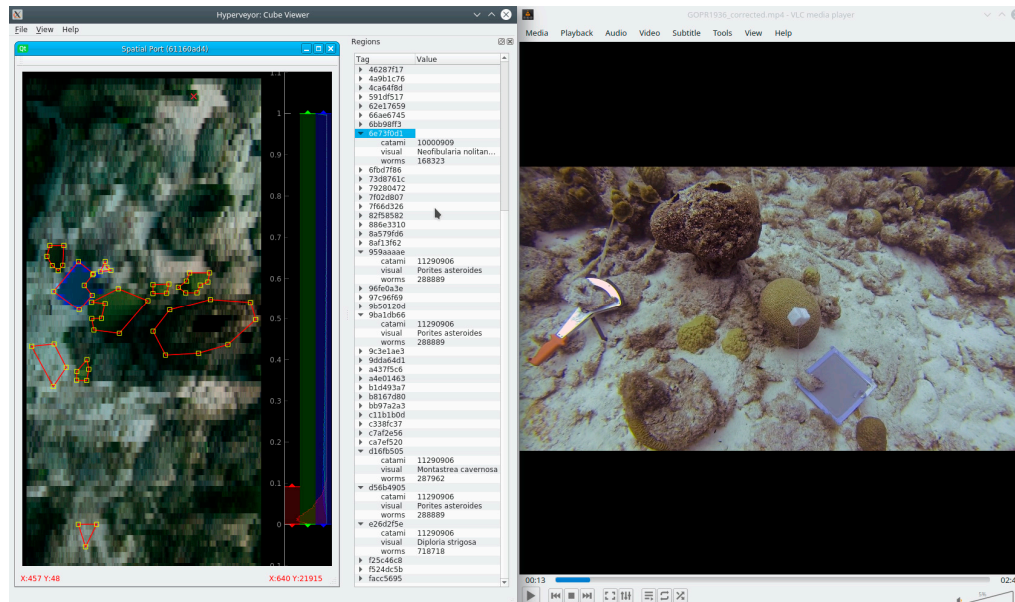


Figure 6. Screen capture showing the annotation of regions on the transect scene (on the left) and accompanying high-resolution video (on the right) to assist in identification. The marked regions are assigned multiple labels, to the deepest taxonomic/hierarchical level possible.

The selection of annotation ROIs was not random. Instead, for each transect, we selected and marked regions to generate the largest number of unique class labels. This was achieved through marking only a few ROIs (3–10) for the most commonly occurring class labels (e.g., “Sediment”, “Turf algae”, etc.) and then focusing on finding new classes to annotate. A given target’s area was marked with several ROI polygons without covering the entire target, thus generating a spatially under-segmented labeling of the target. We also tried to ensure that ROIs were marked across different structural aspects (like shadow areas) of the targets. Labels were assigned to each ROI using the previously described schema (Section 3.3), with the label assigned at the deepest possible hierarchical/taxonomic level, which was variable depending on the target type and image quality at that location.

As each transect’s annotation was approached as an independent evaluation with a focus on maximizing unique classes in the transect scene, the resulting set of labels reflects the class balance distribution of the surveyed area and with a spatial localization that was not random. As a result, there is severe class imbalance across the annotated survey dataset due to classes that appear in every transect (e.g., “Sediment”) and classes that appear in only one or few transects (e.g., *Acropora palmata*). Overall, this approach prioritized expert attention towards greater biodiversity coverage instead of spatial coverage. This greatly reduced annotation effort by avoiding relabeling very common target classes as would be necessary by a random selection of ROIs.

Our curated and annotated dataset is designed to be used in training machine learning classifiers for automated habitat mapping. Our annotation protocol described above was formulated to

maximize two aspects, i.e. the number of classes represented in the annotated dataset, as well as the number of cases of each class. This could enable a well-trained classifier to predict class labels in unseen transects as well as classify benthic taxa, even if they are rare.

4. User Notes

The dataset is organized into transects, with one data folder each (Figure 7). Each transect folder has a JavaScript Object Notation (JSON) file `metadata.json`, containing the name of the surveyed site location and original source data folder. The HyperDiver's hyperspectral image and auxiliary sensor data are saved in a single file `transect.nc`, in Network Common Data Form or netCDF format. The netCDF format can be inspected through a variety of libraries in multiple programming languages, or through a graphical browser such as Panoply (www.giss.nasa.gov/tools/panoply).

The hyperspectral image is saved under the array `cube`, which is a 3-dimensional array with dimensions (y, x, w) . The 8-bit integer `cube` array contains the raw radiance signal in 400 wavelength bands at each spatial location. If local reflectance is required, it can be derived by dividing the raw signals by a reference value obtained from the signal of the gray reference boards (Section 3.1). For each transect, this spectral information is available as ROIs labelled with the `visual` tag "refboard". This will account for differences in solar spectrum across different sites and depths. Using standard red, blue and green channels a natural color rendering of the scene was extracted from the hyperspectral image and saved as a composite image (`natural.jpg`).

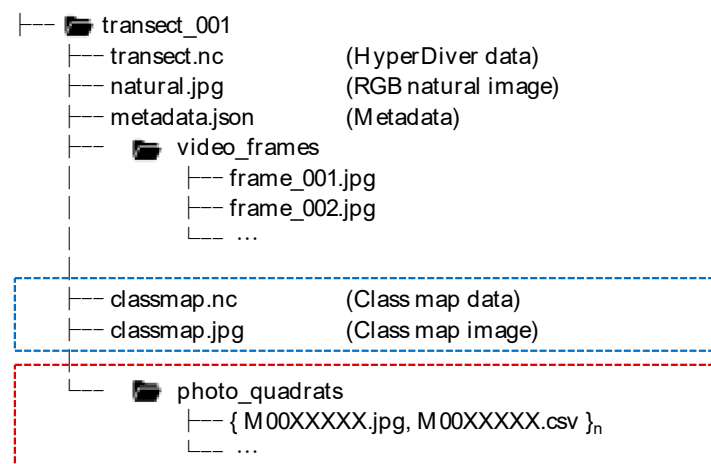


Figure 7. Data are organized into individual folders for each surveyed transect (transect_001 to transect_147). Every data folder contains HyperDiver data, a natural color image as well as extracted video frames. A subset of the transects will also contain additional files containing annotations of hyperspectral data (blue) and/or an additional sub-folder containing photo quadrat images (red).

The other arrays in `transect.nc`, one-dimensional and indexed by timestamp, contain data from auxiliary sensors on the HyperDiver: pressure, altitude, acceleration, pose, and irradiance, each calibrated according to manufacturer specifications. Pressure (in bar) and altitude (in meter) allow localization of the HyperDiver in the water column. Acceleration data from a gyroscope is recorded for three axes, and the pose data such as heading, pitch and roll (in $10 \times$ degree) are found in separate arrays. These data can be used to reconstruct the pose and motion of the HyperDiver during the scan, potentially enabling automated image rectification of each hyperspectral scan. The radiance information (PAR) is saved in units of $\mu\text{mol photons m}^{-2} \text{s}^{-1}$.

All transect scenes are also accompanied by a series of `.jpg` image files of still frames extracted from the corresponding high-resolution video (Figure 6). These are found in a sub-folder named "video_frames" (Figure 7).

Hyperspectral images which have been annotated contain another netCDF file (`classmap.nc`) containing the labels for each ROI as an integer-coded map. While this enables ease of use

downstream in the machine learning workflow, the text labels for the `visual` tag can be recovered from the `netCDF` attributes of each `classmap.nc` file. These integer-coded maps were visualized as a false-color image (`classmap.jpg`) using a global colormap for ease of inspection, with black indicating unlabeled regions. When generating false color habitat maps with trained classifiers, users are reminded to use a similar global label list for color mapping to ensure consistency across all transects, regardless of the set of unique classes from each transect.

Transects that were also surveyed using the photo quadrat method contains a sub-folder called “`photo_quadrats`” that contains all the co-located photo quadrat (`.jpg` files) and annotated labels in a comma-separated values (`.csv`) text file. This text file contains the pixel coordinates and visual labels for all 80 points in the corresponding quadrat image.

Users of this dataset who intend to apply hierarchical learning for classification should note that all class labels of the annotated data are the leaf nodes of the hierarchy labels. To recover the entire hierarchy, users are reminded to reconstruct the tree from two schema, i.e. WoRMS and CATAMI (Section 3.3). First, use the `worms` tag of each ROI label in `classmap.nc` to query the parent nodes of those labels using the WoRMS REST webservice (www.marinespecies.org/rest/) to construct a tree-of-life of biotic labels. The rest of the tree can be constructed based on the CATAMI classification schema and the fully constructed tree diagram published here (Appendix A; Figure A1).

Author Contributions: Conceptualization, A.R.R. and A.C.; methodology, A.R.R. and A.C.; software, A.R.R. and A.C.; validation, A.R.R. and A.C.; formal analysis, A.R.R. and A.C.; investigation, A.C.; resources, A.C.; data curation, A.R.R.; writing—original draft preparation, A.R.R.; writing—review and editing, A.R.R. and A.C.; visualization, A.R.R.; supervision, A.C.; project administration, A.C.; funding acquisition, A.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Max Planck Society.

Acknowledgments: We thank Paul Färber, Harald Osmer and Georg Herz from the workshops of the Max Planck Institute in Bremen for technical support with instrumentation. We thank Joost den Haan, Hannah Brocke, Ben Müller and the CARMABI research station for the logistical support during the field work in Curaçao, and Dirk de Beer for constructive feedback.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

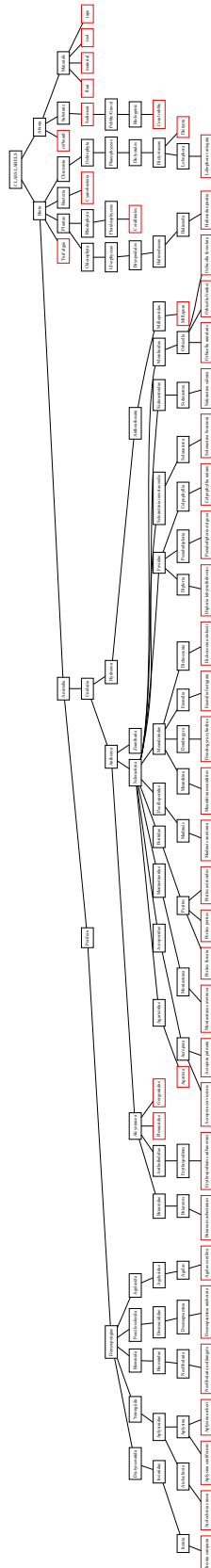


Figure A1. Full extent of the hierarchical labels found in the dataset, best viewed digitally.

References

- Spalding, M.D.; Ravilious, C.; Green, E.P. *World Atlas of Coral Reefs*; University of California Press: Berkeley, CA, USA, 2001; ISBN 978-0-520-23255-6.
- Alcolado, P.M.; Alleng, G.; Bonair, K.; Bone, D.; Buchan, K.; Bush, P.G.; De Meyer, K.; Garcia, J.R.; Garzón-Ferreira, J.; Gayle, P.M.H.; et al. The Caribbean Coastal Marine Productivity Program (CARICOMP). *Bull. Mar. Sci.* **2001**, *69*, 819–829.
- Hodgson, G. Reef Check: The first step in community-based management. *Bull. Mar. Sci.* **2001**, *69*, 861–868.
- Roelfsema, C.M.; Phinn, S.R. Validation. In *Coral Reef Remote Sensing*; Goodman, J.A., Purkis, S.J., Phinn, S.R., Eds.; Springer: Dordrecht, The Netherlands, 2013; pp. 375–401, ISBN 978-90-481-9291-5.
- Maxwell, S.M.; Hazen, E.L.; Lewison, R.L.; Dunn, D.C.; Bailey, H.; Bograd, S.J.; Briscoe, D.K.; Fossette, S.; Hobday, A.J.; Bennett, M.; et al. Dynamic ocean management: Defining and conceptualizing real-time management of the ocean. *Mar. Policy* **2015**, *58*, 42–50.
- Caras, T.; Hedley, J.; Karnieli, A. Implications of sensor design for coral reef detection: Upscaling ground hyperspectral imagery in spatial and spectral scales. *Int. J. Appl. Earth Obs. Geoinf.* **2017**, *63*, 68–77.
- Chennu, A.; Färber, P.; De'ath, G.; de Beer, D.; Fabricius, K.E. A diver-operated hyperspectral imaging and topographic surveying system for automated mapping of benthic habitats. *Sci. Rep.* **2017**, *7*, 7122.
- Dale, L.M.; Thewis, A.; Boudry, C.; Rotar, I.; Dardenne, P.; Baeten, V.; Pierna, J.A.F. Hyperspectral Imaging Applications in Agriculture and Agro-Food Product Quality and Safety Control: A Review. *Appl. Spectrosc. Rev.* **2013**, *48*, 142–159.
- Ghiyamat, A.; Shafri, H.Z.M. A review on hyperspectral remote sensing for homogeneous and heterogeneous forest biodiversity assessment. *Int. J. Remote Sens.* **2010**, *31*, 1837–1856.
- Wentz, E.; Anderson, S.; Fragkias, M.; Netzband, M.; Mesev, V.; Myint, S.; Quattrochi, D.; Rahman, A.; Seto, K. Supporting Global Environmental Change Research: A Review of Trends and Knowledge Gaps in Urban Remote Sensing. *Remote Sens.* **2014**, *6*, 3879–3905.
- Wang, K.; Franklin, S.E.; Guo, X.; Cattet, M. Remote Sensing of Ecology, Biodiversity and Conservation: A Review from the Perspective of Remote Sensing Specialists. *Sensors* **2010**, *10*, 9647–9667.
- Chang, C.-I. *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*; Kluwer Academic/Plenum Publishers: New York, NY, USA; 2003; ISBN 978-0-306-47483-5.
- Mumby, P.J.; Green, E.P.; Edwards, A.J.; Clark, C.D. Coral reef habitat mapping: How much detail can remote sensing provide? *Mar. Biol.* **1997**, *130*, 193–202.
- Pearlman, J.S.; Barry, P.S.; Segal, C.C.; Shepanski, J.; Beiso, D.; Carman, S.L. Hyperion, a space-based imaging spectrometer. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 1160–1173.
- Asner, G.P.; Knapp, D.E.; Boardman, J.; Green, R.O.; Kennedy-Bowdoin, T.; Eastwood, M.; Martin, R.E.; Anderson, C.; Field, C.B. Carnegie Airborne Observatory-2: Increasing science data dimensionality via high-fidelity multi-sensor fusion. *Remote Sens. Environ.* **2012**, *124*, 454–465.
- Foo, S.A.; Asner, G.P. Scaling Up Coral Reef Restoration Using Remote Sensing Technology. *Front. Mar. Sci.* **2019**, *6*, 79.
- Gewali, U.B.; Monteiro, S.T.; Saber, E. Machine learning based hyperspectral image analysis: A survey. *arXiv* **2018**, arXiv:1802.08701.
- Tasar, O.; Tarabalka, Y.; Alliez, P. Incremental Learning for Semantic Segmentation of Large-Scale Remote Sensing Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 3524–3537.
- Greenstein, B.J.; Pandolfi, J.M. Escaping the heat: Range shifts of reef coral taxa in coastal Western Australia. *Glob. Chang. Biol.* **2008**, *14*, 513–528.
- Precht, W.F.; Aronson, R.B. Climate flickers and range shifts of reef corals. *Front. Ecol. Environ.* **2004**, *2*, 307–314.
- WoRMS Editorial Board World Register of Marine Species. Available online: <https://www.marinespecies.org> at VLIZ (accessed on 4 December 2019). DOI: <https://doi.org/10.14284/170>.
- Vandepitte, L.; Vanhoorne, B.; Decock, W.; Dekeyser, S.; Trias Verbeeck, A.; Bovit, L.; Hernandez, F.; Mees, J. How Aphia—The Platform behind Several Online and Taxonomically Oriented Databases—Can Serve Both the Taxonomic Community and the Field of Biodiversity Informatics. *J. Mar. Sci. Eng.* **2015**, *3*, 1448–1473.
- CATAMI Technical Working Group. *CATAMI Classification Scheme for Scoring Marine Biota and Sub-Strata in Underwater Imagery. Version 1.4*; National Environmental Research Program, Marine Biodiversity Hub: Canberra Australia, 2014.

24. Pante, E.; Dustan, P. Getting to the Point: Accuracy of Point Count in Monitoring Ecosystem Change. *J. Mar. Biol.* **2012**, *2012*, 1–7.
25. Kohler, K.E.; Gill, S.M. Coral Point Count with Excel extensions (CPCe): A Visual Basic program for the determination of coral and substrate coverage using random point count methodology. *Comput. Geosci.* **2006**, *32*, 1259–1269.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).