Scaling the Annotation of Subtidal Marine Habitats

Perry Naughton[1], Clinton Edwards[2], Vid Petrovic[1], Ryan Kastner[3], Falko Kuester[1], and Stuart Sandin[2]

[1]Center of Interdisciplinary Science of Art, Architecture and Archaeology, University of California, San Diego

[2]Center for Marine Biodiversity and Conservation, Scripps Institute of Oceanography, University of California, San Diego

[3]Department of Computer Science and Engineering, University of California, San Diego

[*]email: {pnaughto, clint, vipetrov, kastner, fkuester, ssandin}@ucsd.edu

ABSTRACT

Visually documenting seafloor habitats has the potential to answer challenging questions in several maritime disciplines including: ecology, geology, and archaeology. Unfortunately, the attenuation of visible light underwater limits the imaging footprint of a single image to square meters. This limitation makes representing large habitats, on the order of hundreds of square meters and beyond, an intensive process requiring the collection, storage, processing and annotation of thousands of high resolution images per hundred square meters of seafloor. This paper describes a pipeline for dealing with these challenges efficiently and effectively using visual data of coral reef communities processed into a three dimensional model. We evaluate the resources and technological advancements required to scale this problem to orders of magnitude larger than the current state of the art and motivate the need for networked underwater data collection platforms to push the scalability of this method.

1. INTRODUCTION

For many scientific disciplines, image data sets are collected to quantify spatial distributions and dynamic evolution of patterns. For example, aerial and satellite-derived imagery provide world-changing advances in our understanding of terrestrial systems. Comparable large-scale, temporally replicated, and high-resolution data sources are not available for subtidal marine habitats. While single images of the seafloor have long been used to provide archival sources of information, it is difficult to understand the entire context of an area without reconstructing the relationship between images. For this reason, photomosaics and 3D point clouds are increasingly used in domain sciences as succinct representations of image data sets that are intuitive for the end user to use. While we have come a long way in building algorithms and methods to construct these representations, maritime disciplines are still limited by the scale of the data we can collect, process and annotate.

The adoption of image processing techniques has shown increased efficiency in maritime fields such as archeology and ecology. In archeology, shipwrecks have been recreated using images and hand measurements to create an accurate and comprehensive representation in 3D [15, 17, 11, 16, 4]. These photogrammetric techniques have the potential to replace many of the tedious hand drawings (which must be performed underwater) that provide archaeologists with a comprehensive, to scale, visual representation of shipwrecks. Photogrammetric models have also been used to understand marine environments, and recently several authors have demonstrated the utility of these techniques for habitat classifications, largely in the quantification of substrate rugosity [23, 6, 10, 12, 7]. Fewer authors have used such methods in regular ecological monitoring and investigation of key ecological processes [13, 8, 20, 5]. Despite this, given the scale and spatially explicit nature of organismal interactions on coral reefs, there is a critical need to expand these technologies to facilitate ecological inquiry.

Despite the potential of these technologies to enable groundbreaking insights in coral reef ecology and other maritime fields, data collection and analyses have been confined to a modest scale. As a result, our inferences have been constrained to processes acting at that scale. Comparisons among reefs, or within a given reef over time, are made using composite metrics which cannot detect important demographic differences in coral communities. To overcome these limitations, our group has recently begun to implement photometric technology which offers a dramatic increase in the spatial scale of imagery without increasing the amount of underwater time required. Once digitized, photometric models become spatially explicit maps of the benthos which allows us to build on past efforts and open up novel lines of inquiry such as: How do changes in the wave environment change the structure and composition of reef communities? Are there consistent patterns of co-occurrence in benthic organisms on coral reefs? Do corals exhibit trade offs in life history and growth strategies? Human effects? In order to scale up our efforts to the whole island scale, and make meaningful comparisons between islands, we are in need of key innovations that will allow us to increase the scale of our investigation.

Our efforts to find answers to some of these large ecological questions have inspired two research initiatives that we plan to undertake over the next couple of years. The first, coined the '100 Island Challenge', is focused on elucidating spatial patterns across a variety of environmental

WUWNet'15 , October 22-24 2015, Washington DC, USA

and anthropogenic contexts. Over the course of three years, image datasets will be collected from 6-8 replicated $100m^2$ plots from 100 different coral islands in regions around the world. Each island will be visited twice to compare evolving dynamics of spatial patterns to help unravel the effects of different regional context's contributions to differences in observed patterns and processes. This comparative approach will unlock answers to many community ecology questions. In addition to the direct goals of the 100 island challenge, we plan to leverage long term collaborations with governmental and other monitoring agencies, such as the national oceanographic and atmospheric administration's coral reef ecosystem division, to encourage the use of our methodology and approach to many more regional contexts and eventually reach a greater global representation.

The second challenge, dubbed the 'Vostok Challenge', is an ambitious interdisciplinary project to map the entire island of Vostok, Republic of Kiribati. A nearly exhaustive sampling design focused on imaging from the smallest (molecule) to the largest (island) scale will allow us to produce the most complete coral reef dataset ever, representing an almost limitless platform for ecological inquiry. The extreme isolation of Vostok presents an almost unprecedented opportunity for real time communication of expedition progress and live data feeds, with few comparisons beyond the space program. These two initiatives will expand our knowledge about coral reef communities and push engineering capacity to collect, store and manage this amount of data. Ultimately, we hope to combine the '100 Island Challenge' framework for island comparisons with the technological advancements achieved during the 'Vostok Challenge' to create the capability to process the next order of magnitude - mapping the entirety of 100 islands.

Our current pipeline requires high end computing as well as significant human data collection and annotation. The end result is typically a three dimensional model of coral reefs covering a 10m by 10m footprint. Starting with the data collection, it takes a team of two divers 1 hour to gather the 2500 images. It takes about 2 days to process the images into a 500 million point representation on a high end computer (2 Xeon Processors, 128Gb RAM, 4 GeForce GTX 980 GPUs). The annotation takes 60 hours on a two dimensional representation of the 3D geometry of the $100m^2$ reef with up to 250+ different species to classify, depending on the location. Given our current pipeline, estimates of the person hours and processing time is summarized in Table 1 for different orders of magnitude.

Table 1: Table of Effort per Scale

	Collection (man Hr)	Processing (Compute Hr)	Annotation (man Hr)
10x10m plot	1	50	60
100 Islands	8	400	480
Vostok (250,000 sqm)	2,500	125,000	150,000
100 Islands	250,000	1,250,000	1,500,00

Our data pipeline has allowed us to collect unique data sets, pushing our understanding of coral ecology past what was previously possible. Yet Table 1 points out the limitations and inconveniences of this method. Even for a 10m by 10m plot there is significant computational and human resources that goes into a finished representation. As we consider larger datasets, it is clear that our current pipeline represents a critical bottleneck that must be overcome.

Networked vehicles are key to pushing this method to the next order of magnitude by significantly reducing the time required for the first two stages of the data pipeline. In addition to the added benefit of switching from 'person power' to 'compute power' in the data collection stage (which has more endurance per day and can be less costly), networked vehicles can evaluate their coverage and coordinate trajectories during data collection more effectively than their human counterparts. This will increase the efficiency of the data collection process. Also, vehicles can provide accurate estimates of the pose and orientation as metadata which can increase the efficiency and accuracy of the technical post processing. Importantly, networked vehicles will free us of the limitations of SCUBA, allowing greater spatial coverage and will be able to access a wider (eg deeper) range of habitats without the risks associated with divers operating on SCUBA.

The remainder of this paper is focused on first describing our image pipeline for analyzing coral reefs followed by a description of technological advancements that are needed to scale these methods to a point where they are more effective.

2. CURRENT METHODOLOGY

Our data processing work flow involves three major components. The first component is data collection in the field, followed by technical post-processing of raw imagery into a complete photometric model. Models are then ecologically post-processed by expert biologists to derive information that is useful for coral biology.

2.1 Data Collection

The collection of overlapping digital images is a straightforward process that requires little special equipment or dive operations[13]. The camera system consists of two SLR Nikon D7000 cameras and a single GoPro video camera mounted to a custom frame (Figure 1 A). One camera uses a wide-angle lens to ensure high overlap among adjacent images. The other camera uses a longer focal length lens to capture images with sub-cm spatial resolution. The extreme wide angle GoPro video camera is used in the rare event of missing imagery for a given portion of the reef. Images are captured every second from both cameras simultaneously, with mounted lasers to provide scale (Figure 1 B). The diver operating the camera system swims a double lawnmower (i.e. zigzag, see Figure 1 C) pattern at a speed sufficient to maintain maximum overlap between adjacent images to obtain continuous coverage of the reef floor in a plot of up to $100m^2$ in a single dive. Depending on local conditions, a single data set will take 30-45 minutes to collect and consist of approx. 1200-2500 individual images per camera. A detailed series of measurements is taken between fixed markers visible in the imagery to aid in image calibration (see highlights in Figure 1 D).

2.2 Data Processing

Our current data processing stage employs a technique called Structure from Motion (SfM). SfM compares multiple camera views to estimate the 3D geometry of an object in the form of a point cloud. Each point in a point cloud is generated from pixels in an image. At a high level, SfM

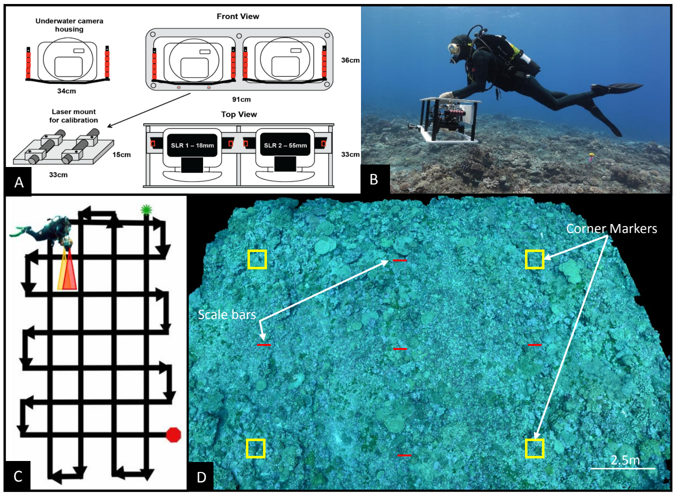

Figure 1: (A) Schematic of the diver rig with cameras of different focal lengths and laser pointers for scale (B) Photo of a diver swimming transects (C) representation of the path swam by the diver to insure adequate coverage (D) end result of the photometric model with markers highlighted

starts by picking out thousands of salient points in each image. It then finds matches of salient points between images, determining which images have overlap. With information of point matches the software estimates the relative position and orientation of each frame. Finally, the 3D location of the points in the image can be deduced through triangulation and refined by a process called bundle adjustment. In general, SfM provides a geometrically accurate representation of a scene, with the caveat that there is no scale. When using a single camera, it is ambiguous whether a camera is moving a small distance around a small object or a large distance around a large object. For this reason, some scale must be defined by using known distances or dimensions of one or more objects. In other words, physical measurements of an object are still crucial to provide a correct scale. There are multiple implementations of SfM software, both free (VisualSfM [22], Autodesks 123D Catch[1], Bundler[21]) and for sale (Agisoft[14]). In this study we chose to use Agisoft because we have found that it is the most comprehensive.

In order to define a scale we insert ground control points (GCP) in the scene we are focused on so that we have known distances in our model. In our case we have rebar markers in the coral structure. To establish coordinates we measure the distances between the markers with a tape measure and collect the depth at the marker with a pressure gauge. With this information we can solve for a relative location of the markers that minimizes the discrepancy of each measurement in the least squares sense. Once a datum is tagged in the imagery, the coordinates for that datum can be entered, defining the scale of the final SfM model and helping the reconstruction accuracy.

2.3 Ecological Post-Processing

Our interest in this whole method is to create taxonomically classified maps of the seafloor so that detailed analyses of pattern and process can be undertaken. Therefore, a critical step in the extraction of information lies in ecological post-processing where skilled biologists outline and taxonomically identify all individual benthic organisms by hand. The end result are accurate maps of community structures between taxa. While this classification step is a time intensive process (requiring approximately 60 man hours), it is tremendously more efficient and comprehensive than in situ methodologies which are limited by a diver's bottom time. In contrast, photometric models allow for limitless 'bottom-time' for exhaustive searches of large areas of the benthos in a relatively rapid manner.

The goal is to perform classification and annotation interactively on the derived 3D model with convenient access to the information needed to do the task reliably. Our basic annotation paradigm is that of painting: directly marking or highlighting regions using brushes with a dynamically-controlled radius. We choose to work directly on the dense pointcloud, since it is the most detailed reconstructed product, from which we can construct other (further filtered /interpolated) models (such as meshes, DEMs, and orthophotos) as needed, and assign annotations per-point, partitioning the point cloud according to the classification. Note also that the derived products of the photometric reconstruction techniques outlined in the preceding section include not only an approximate 3D model of the imaged scene, but also pose and intrinsic camera parameter estimates for the input images. These estimates permit the source images - together comprising the primary captured record of the site - to be consulted efficiently as a part of the 3D classification workflow. In particular, we can automatically look up on-demand and overlay each relevant source image on the pointcloud, or display it in a separate window as an aid in performing the classification.

There are some challenges in building software to view and annotate such 3D representations. The first is dealing with the size of the 3D datasets involved, which can easily exceed 500 million points per study region. To visualize the pointclouds, we apply the point-buffering approach of [19], and extend it by reordering the points to accelerate spatial queries. For each point set, we construct a balanced 3D tree [3] and sort the points within the set such that the tree structure is implicitly encoded by the point order, thereby allowing us to store the trees with very low overhead (0.4% on average).

The second challenge is designing and implementing the tools needed to support an efficient classification and annotation workflow. To this end, we have built into our system a run-time scripting framework (currently using the V8 JavaScript engine [9]), allowing us to rapidly implement the front-end for the annotation workflow, and providing greater flexibility for developing and testing a variety of annotation tools. Within the scripting environment, we expose spatial queries (eg, nearest-neighbor - 'the set of points within a certain distance of a given point', accelerated by the above mentioned implicitly-stored 3D trees) that serve as building blocks for more elaborate local geometric operations (such as flood-fill style 'magic' brushes).

We implement the remainder of the front-end in a web-app (that can run in a separate browser window, or on a different machine or mobile device) providing control over annotations, allowing the list of classes to be edited, metadata to be updated, and display parameters to be set (e.g. which classes are shown, and in what color). Finally, we leverage the scripting framework to develop analytical tools (measurements, statistics) as well as a flexible means of exporting data for further processing (separating points by class, subsampling /filtering to prepare a point set for meshing).

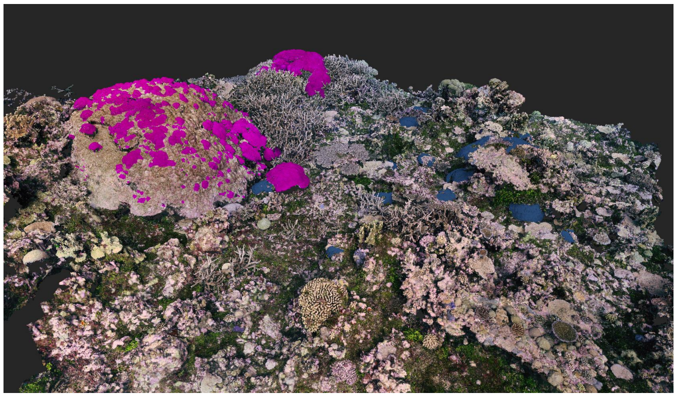

Figure 2: A 3D model output from SfM of Ant Atoll. This model represents a typical 10m by 10m plot of coral. The different colors represent different annotations that we have our software showing.

3. FUTURE DEVELOPMENTS

Our current method for digitally analyzing coral reef communities is an intensive process. Both the number of man hours and the compute time can become infeasible for large datasets. As we consider larger scales we will inevitably turn towards vehicles for the data collection process to extend the endurance and efficiency of our data collection. As an added benefit, vehicles also have the potential to increase the accuracy and decrease the processing time by annotating each image with an estimate of the pose and orientation.

Utilizing a network of vehicles can also improve the efficiency and accuracy of the data collection process past being able to parallelize the collection. In past work we have shown that vehicle location estimates can be jointly improved by collaboratively localizing and communicating position estimates between vehicles [18]. The increase in localization accuracy will allow higher confidence in our coverage and will allow less time to be spent collecting redundant data.

3.1 Data Collection

Currently there is no feedback on our data collection process. This requires us to be conservative in our data collection and results in redundant coverage. For example, a minimum of 3 images must include a point in order for the point to be resolved by SfM software. More than 3 images can help with the accuracy of the 3D point estimate but the additional gain in accuracy quickly deteriorates with the number of images. There is a trade off between how much one more image will improve the accuracy and the amount of time it takes to capture that image. In our data sets we are seeing coverage in excess of 50 images per point. While we have not determined the optimal coverage for our models, we believe that most of the information is redundant. Once we better understand the optimal coverage of images per point we can design our vehicle surveys to target this coverage. This is possible because vehicles have an estimate of their position during data collection and can provide an estimate of how many images overlap with each point on the ocean floor.

This problem will be exacerbated once we scale to larger data sets, such as a whole island where we need to consider the overlap between deployments as well. In Table 1, we assumed that we could map disjoint 10x10m areas and stitch them together, but this is not true. We will need the overlap both for the accuracy of the model and to insure that all of the area is accounted for. If we do not know our position estimates with confidence then we will be forced to be conservative in the data collection to insure that we collect everything that we need. Using vehicles with accurate position estimates will greatly alleviate this process yet it is clear that many scales will still be ambitious. We believe that the scale of the full island will still be time consuming for the current state of the art of networked vehicles.

3.2 Data Processing

A significant part of the processing time for SfM is determining which images have overlap. In the naive case this step scales quadratically with the number of input images, as we check each image with every other image for overlap. With some of our larger datasets we have seen that this step can take over 40% of the processing time. With the localization estimates that vehicles provide we will be able to determine which images have overlap a priori and we can reduce the processing time by almost half. Additionally, utilizing different sensors will alleviate the need to define a geometry in our photometric models, which is currently the only part which requires human intervention. Since each image will already have metadata of location to a real world scale, we will be able to utilize this data to accurately scale these models. Likewise, if two networked vehicles collect disjoint image sets, we will be able to construct the relative position between the photometric models without knowledge of the missing data.

3.3 Ecological Post Processing

The most daunting part of our pipeline moving forward is the annotation of different coral species in each of the models we create. The ecological analysis requires high classification rates of hundreds of different coral species. Stringent annotation requirements are needed to detect subtleties in community structure of the coral ecosystems. Recent advancements in computer aided classification may soon expedite this process. Due to the complexity of coral reef habitats and the *lack of existing training data sets*, accuracy levels of existing technologies are still insufficiently low [2]. After enough time spent on the annotation we will eventually have a comprehensive data set that we can use to generate classification algorithms to do most of the classification automatically. We have done some preliminary studies where we have achieved >90% classification accuracy for certain subsets of locations and species. Eventually we can get to a point where we are simply checking the output of the classification algorithms instead of doing the whole classification ourselves. This will significantly speed up our classification throughput on the classification side and allow us to understand ecological patterns across large swaths of data.

4. CONCLUSION

We are working on understanding coral reef and other maritime habitats on a scale that has not been studied before. The methods to work on this scale are intensive in compute time and in manual labor from SCUBA divers and biologists. However, these methods are orders of magnitude more comprehensive than performing the analysis in situ. There is great opportunity for increased efficiency and much of this efficiency will come from utilizing underwater vehicle platforms that work in teams to increase localization accuracy and efficiently collect data. These advances are much needed for measuring human impacts, identifying key metrics of healthy sub-tidal habitats and constructing conservation plans to keep oceans healthy. An additional benefit of this work will arise through the volumes of human classified data and subsequent innovations in automated annotations. The progress made in this sector will likely have far reaching commercial applications in the fields of remote environmental sensing and robotics.

5. ACKNOWLEDGMENTS

This work was supported by the National Science Foundation under IGERT Award #DGE-0966375, "Training, Research and Education in Engineering for Cultural Heritage Diagnostics." Additional support was provided by the Gordon and Betty Moore foundation reefs of tomorrow initiative Qualcomm Institute at UC San Diego, the Friends of CISA3 and the World Cultural Heritage Society. Opinions, findings, and conclusions from this study are those of the authors and do not necessarily reflect the opinions of the research sponsors.

6. REFERENCES

[1] Autodesk. 123d catch. `http://www.123dapp.com/catch`, 2015.

[2] O. Beijbom, P. J. Edmunds, D. Kline, B. G. Mitchell, D. Kriegman, et al. Automated annotation of coral reef survey images. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 1170–1177. IEEE, 2012.

[3] J. L. Bentley. Multidimensional binary search trees used for associative searching. *Communications of the ACM*, 18(9):509–517, 1975.

[4] P. Bojakowski, K. Bojakowski, and P. Naughton. A comparison between structure from motion and direct survey methodologies on the warwick. *Journal of Maritime Archaeology*, 10(2):159–180, 2015.

[5] J. Burns, D. Delparte, R. Gates, and M. Takabayashi. Integrating structure-from-motion photogrammetry with geospatial software as a novel technique for quantifying 3d ecological characteristics of coral reefs. *PeerJ*, 3:e1077, 2015.

[6] P. Dustan, O. Doherty, and S. Pardede. Digital reef rugosity estimates coral reef habitat complexity. *PloS one*, 8(2):e57386, 2013.

[7] A. Friedman, O. Pizarro, S. B. Williams, and M. Johnson-Roberson. Multi-scale measures of rugosity, slope and aspect from benthic stereo image reconstructions. 2012.

[8] A. C. Gleason, D. Lirman, D. Williams, N. R. Gracias, B. E. Gintert, H. Madjidi, R. Pamela Reid, G. Chris Boynton, S. Negahdaripour, M. Miller, et al. Documenting hurricane impacts on coral reefs using two-dimensional video-mosaic technology. *Marine Ecology*, 28(2):254–258, 2007.

[9] Google. v8 - v8 javascript engine - google project hosting. `https://code.google.com/p/v8/`, September 2015.

[10] H. He, R. Ferrari, D. McKinnon, G. Roff, P. Mumby, R. Smith, and B. Upcroft. Measuring reef complexity and rugosity from monocular video bathymetric reconstruction. In *Proceedings of the 12th International Coral Reef Symposium*, pages 1–5, 2012.

[11] J. Henderson, O. Pizarro, M. Johnson-Roberson, and I. Mahon. Mapping submerged archaeological sites using stereo-vision photogrammetry. *International Journal of Nautical Archaeology*, 42(2):243–256, 2013.

[12] J. Leon, C. M. Roelfsema, M. I. Saunders, and S. R. Phinn. Measuring coral reef terrain roughness using 'structure-from-motion' close-range photogrammetry. *Geomorphology*, 242:21 – 28, 2015.

[13] D. Lirman, N. Gracias, B. Gintert, A. Gleason, R. Reid, S. Negahdaripour, and P. Kramer. Development and application of a video-mosaic survey technology to document the status of coral reef communities. *Environmental Monitoring and Assessment*, 125(1-3):59–73, 2007.

[14] A. LLC. Agisoft photoscan. `http://www.agisoft.com/`, 2015.

[15] I. Mahon, S. Williams, O. Pizarro, and M. Johnson-Roberson. Efficient view-based slam using visual loop closures. *Robotics, IEEE Transactions on*, 24(5):1002–1014, Oct 2008.

[16] J. McCarthy and J. Benjamin. Multi-image photogrammetry for underwater archaeological site recording: An accessible, diver-based approach. *Journal of Maritime Archaeology*, 9(1):95–114, 2014.

[17] J. Mertes, T. Thomsen, and J. Gulley. Evaluation of structure from motion software to create 3d models of late nineteenth century great lakes shipwrecks using archived diver-acquired video surveys. *Journal of Maritime Archaeology*, 9(2):173–189, 2014.

[18] D. Mirza, P. Naughton, C. Schurgers, and R. Kastner. Real-time collaborative tracking for underwater networked systems. *Ad Hoc Networks*, pages –, 2014.

[19] V. Petrovic, D. J. Vanoni, A. M. Richter, T. E. Levy, and F. Kuester. Visualizing high resolution three-dimensional and two-dimensional data of cultural heritage sites. *Mediterranean Archaeology and Archaeometry*, 14(4):93–100, 2014.

[20] R. P. Reid, D. Lirman, N. Gracias, S. Negahdaripour, A. Gleason, and B. Gintert. Application of landscape mosaic technology to complement coral reef resource mapping and monitoring. Technical report, DTIC Document, 2010.

[21] N. Snavely, S. M. Seitz, and R. Szeliski. Photo tourism: exploring photo collections in 3d. In *ACM transactions on graphics (TOG)*, volume 25, pages 835–846. ACM, 2006.

[22] C. Wu. Visualsfm: A visual structure from motion system. `http://ccwu.me/vsfm/`, 2015.

[23] P. Zapata-Ramírez, D. Scaradozzi, L. Sorbi, M. Palma, U. Pantaleo, M. Ponti, and C. Cerrano. Innovative study methods for the mediterranean coralligenous habitats. *Advances in Oceanography and Limnology*, 4(2):102–119, 2013.