Journal of Visualized Experiments

A Field Primer for Monitoring Benthic Ecosystems Using Structure-from-Motion Photogrammetry

--Manuscript Draft--

Article Type:	Invited Methods Article - JoVE Produced Video
Manuscript Number:	JoVE61815R1
Full Title:	A Field Primer for Monitoring Benthic Ecosystems Using Structure-from-Motion Photogrammetry
Corresponding Author:	Ty Noble Frederick Roach, PhD University of Hawai'i at Manoa Hawaii Institute of Marine Biology Kāneʻohe, HI UNITED STATES
Corresponding Author's Institution:	University of Hawai'i at Manoa Hawaii Institute of Marine Biology
Corresponding Author E-Mail:	smokinroachjr@gmail.com
Order of Authors:	Ty Noble Frederick Roach, PhD
	Spencer Miller
	Shreya Yadav
	Carlo Caruso
	Jenna Dilworth
	Catherine M Foley
	Joshua R Hancock
	Joel Huckeba
	Ariana S Huffmyer
	Kira Hughes
	Valerie A Kahkejian
	Elizabeth M.P. Madin
	Shayle B Matsuda
	Michael McWilliam
	Erika P Santoro
	Mariana Rocha de Souza
	Damaris Torres-Pullizaa
	Crawford Drury
	Joshua S Madin
Additional Information:	
Question	**Response**
Please indicate whether this article will be Standard Access or Open Access.	Open Access (US$4,200)
Please indicate the **city, state/province, and country** where this article will be **filmed**. Please do not use abbreviations.	Honolulu, HI
Please confirm that you have read and agree to the terms and conditions of the	I agree to the Author License Agreement

author license agreement that applies below:	
Please specify the section of the submitted manuscript.	Environment
Please provide any comments to the journal here.	Thank you for your time and consideration.

Title:

A Field Primer for Monitoring Benthic Ecosystems Using Structure-from-Motion Photogrammetry

Authors:

Ty N.F. Roach[1], Shreya Yadav[1], Carlo Caruso[1], Jenna Dilworth[1,2], Catherine M. Foley[1], Joshua R. Hancock[1], Joel Huckeba[1], Ariana S. Huffmyer[1], Kira Hughes[1], Valerie A. Kahkejian[1], Elizabeth M. P. Madin[1], Shayle B. Matsuda[1], Michael McWilliam[1], Spencer Miller[1,3], Erika P. Santoro[4], Mariana Rocha de Souza[1], Damaris Torres-Pulliza[1], Crawford Drury[1], Joshua S. Madin[1]

[1]Hawai'i Institute of Marine Biology, University of Hawai'i Manoa,Kāne'ohe, HI, USA
[2]Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, FL, USA
[3]Hawai'i Pacific University, Honolulu, HI, USA
[4]Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

Shreya Yadav: shreyay@hawaii.edu
Carlo Caruso: carloxcaruso@gmail.com
Jenna Dilworth: jenna.dilworth@rsmas.miami.edu
Catherine M. Foley: cmrfoley@gmail.com
Joshua R. Hancock: hancock9@hawaii.edu
Joel Huckeba: joel_huckeba@sbcglobal.net
Ariana S. Huffmyer: ashuff@hawaii.edu
Kira Hughes: kirawa@hawaii.edu
Valerie A. Kahkejian: vkahkejian@gmail.com
Elizabeth M.P. Madin: emadin@hawaii.edu
Shayle B. Matsuda: shayle@hawaii.edu
Michael McWilliam: mjmcwilliam@outlook.com
Spencer Miller: smiller11@my.hpu.edu
Erika P. Santoro: erikapsantoro@gmail.com
Mariana Rocha de Souza: mrds@hawaii.edu
Damaris Torres-Pulliza: dama.reefs@gmail.com
Crawford Drury: druryc@hawaii.edu
Joshua S. Madin: jmadin@hawaii.edu

Corresponding Author:
Ty Roach
smokinroachjr@gmail.com

Key Words:

SfM photogrammetry, Field methods, Benthic surveys, Ecological monitoring, Community composition, Habitat structure,

Summary:

We provide a detailed protocol for conducting underwater structure-from-motion photogrammetry surveys to generate 3D models and orthomosaics.

Abstract:

Structure-from-motion (SfM) photogrammetry is a technique used to generate three-dimensional (3D) reconstructions from a sequence of two-dimensional (2D) images. SfM methods are becoming increasingly popular as a noninvasive way to monitor many systems, including anthropogenic and natural landscapes, geologic structures, and both terrestrial and aquatic ecosystems. Here, a detailed protocol is provided for collecting SfM imagery to generate 3D models of benthic habitats. Additionally, the cost, time efficiency, and output quality of employing a Digital Single Lens Reflex (DSLR) camera versus a less expensive action camera have been compared. A tradeoff between computational time and resolution was observed, with the DSLR camera producing models with more than twice the resolution, but taking approximately 1.4-times longer to produce than the action camera. This primer aims to provide a thorough description of the steps necessary to collect SfM data in benthic habitats for those who are unfamiliar with the technique as well as for those already using similar methods.

Introduction

Ecosystem processes are naturally dynamic and can be difficult to quantify. The past decade has seen a surge in new technologies for capturing ecosystems and their dynamics in a range of scales from 3D laser scanning of individual ecosystem features to satellite remote sensing of large areas[1–3]. In benthic habitats, structure is intimately connected with ecosystem function[8], making tools that simultaneously allow for monitoring geometry and community structure especially valuable for understanding ecological dynamics. However, many modern approaches cannot be used in aquatic systems due to the physical properties of water (e.g., refraction, distortion, turbidity). Techniques, such as Light Detection and Ranging and some aerial survey methods, may be appropriate on large spatial scales, but cannot acquire the resolution needed to assess fine scale changes in benthic habitats. Structure-from-Motion (SfM) photogrammetry methods have recently been adapted to produce large-scale, high-resolution orthomosaics and 3D surface models of underwater habitats[4–7].

SfM photogrammetry is a relatively low-cost, simple, non-invasive, and repeatable method that allows for the generation of large-scale, high-resolution records of the benthic environment in aquatic ecosystems[9]. SfM uses a sequence of 2D images to generate 3D model reconstructions. The models generated from SfM can be used to collect data on the structural complexity (e.g., rugosity, dimensionality)[4,5,10–12] and community structure (e.g., species composition, population demography)[13–15] of benthic ecosystems. Furthermore, as this method is relatively inexpensive, quick, and repeatable, it can be used by both scientists and non-scientists to gather valuable, objective information on these ecosystems. Therefore, this method is a viable technique for use in citizen science projects where standardization of sampling effort, minimization of bias, engagement of participants, and ease of training are vital to the quality of data and overall success[16,17].

This article provides a detailed protocol for conducting underwater SfM surveys. Simultaneously, the use of a DSLR camera has been compared with that of a more cost-effective ‘action camera’, and the relative advantages and disadvantages of each are outlined. The overall objective is to familiarize scientists and non-scientists with benthic SfM survey methods as rapidly as possible by providing a simple, commonly used protocol, in turn, promoting the use of this method more widely. Examples of studies that have applied variations of this method to study underwater ecological communities have been published[4,10,14,18,19–22].

The method described here requires a two-person snorkel or SCUBA team. After the survey site is selected, a spool of line (**Figure 1A**) is placed at the center of the site, and calibration tiles (**Figure 1B**) are distributed ~2 m from the center. One person (the swimmer) swims with the camera and captures images of the site, while the second person (the assistant) tends the spool in the center of the plot (**Figure 1C**). First, the swimmer connects the camera to the spool via the line and then begins to take continuous pictures of the benthos while swimming face-down and forward to unwind the line off the spool. The swimmer should maintain a vertical distance of ~1 m above the substrate at all times, adjusting their position to match that of the topography as they swim. Importantly, the line connecting the spool and camera should remain taut at all times to create even spacing in the spiral as the swimmer surveys the plot. The assistant maintains the spool in a stable, upright position and ensures that the spool does not rotate, and that the line does not become tangled.

Once the line has been completely unwound, the swimmer stops, turns, and swims in the opposite direction to recoil the line around the spool. As the swimmer switches directions, the assistant turns the spool to wind the line in, exactly 180° to prevent exact overlap of the outgoing path. Once the swimmer is as close to the center as possible, the camera is detached from the line, and the assistant takes the spool and line and swims away from the central portion of the site. The swimmer then finishes imaging the center of the plot by moving the camera in a small spiral over the center. While there are several ways to image an area effectively, the spool-and-line method described here is robust in even non-ideal environmental conditions where choppy surface waters, swell, or low visibility might otherwise impede data collection. In these scenarios, this method keeps snorkelers/divers attached and ensures high overlap of images by keeping the swimmer on a controlled path.

Protocol:

1. **Materials**

1.1. Camera

1.1.1. Ensure minimum specifications of durability and waterproof nature (or a waterproof housing) and a minimum frame rate of 2 frames/s (fps).

NOTE: A minimum frame rate of ~4 fps was used in this example.

1.1.2. Digital Single Lens Reflex (DSLR) camera

1.1.2.1. Set the camera to shoot continuously at a photo capture rate between 2 fps and 5 fps.

1.1.2.2. To reproduce the protocol described for this example, use a camera in an underwater housing (see **Table of Materials**) with the following settings: **Manual Mode (M); f10, 18 mm; shutter speed = 1/320; exposure compensation = -1/3; image quality = highest, no RAW; drive mode = continuous; autofocus = AI SERVO; ISO = Auto, max3200; file numbering = Auto reset; image auto rotate = Off; time/date = UTC**.

1.1.3. Action camera

1.1.3.1. Set to video mode or continuous shooting mode at the highest resolution and frame rate possible.

NOTE: The action camera can also be used in continuous mode as long as the frame rate is 2 images per second or greater.

1.1.3.2. To reproduce the protocol in this example (see **Table of Materials**), use a waterproof action camera with the following settings: Video resolution = 4K (4:3 aspect ratio); frame rate = 30 fps.

NOTE: For action cameras, it may be easier to attach the line from the spool to the swimmer rather than to the camera. In this example, the line was attached to the swimmer's wrist via a small lanyard.

1.2. Spool rig (**Figure 1A**)

1.2.1. Ensure that the spool is of the appropriate size to hold the length of line needed for the survey site radius.

NOTE: The circumference of the spool controls the spacing of the spiral swim lines, and the length of the line determines the sample area. In this example, an ~8 inch (~20 cm) diameter spool was used for ~50 inch (~1.3 m) spacing of swim lines. See [9] for details.

1.2.2. Select a spool rig with a flanged edge (for smoothly guiding the line on and off the spool) and attachment points for a handle and pole (to control height from substrate). Ensure that the spool rig is inherently negatively buoyant or made so with the addition of weights.

NOTE: In this example, polyvinyl chloride (PVC) pipes for the handle and pole were used, and the spool was 3D printed in polylactic acid plastic. However, the spool can be as simple as a large PVC pipe or any other round object with the desired diameter.

1.2.2.1. For frequent use and/or challenging field conditions, select a spool made of a more rugged material such as aluminum.

1.2.2.2. Make sure that the spool does not rotate on the pole or spin when in use.

1.2.3. Fix the line to the spool at one end and to a detachable clip at the other for connecting to the camera.

NOTE: The length of the line defines the radius of the site. Here, 6 m of line was used for sites of 12 m in diameter.

1.3. Calibration tiles

1.3.1. Although specialized calibration tiles are not necessary, ensure that negatively buoyant, recognizable objects of known size are included in the model for scale. Consider surge and current conditions to ensure that suitable materials are used so that tiles remain stationary during photo collection.

NOTE: Here, scale marker templates available as part of certain software programs were printed on waterproof paper, which was attached to 1-inch-thick PVC tiles.

1.3.2. Provide divers with a means to measure the depth of the tile such as an electronic depth gauge (see **Table of Materials**).

1.4. Color correction

1.4.1. Set white balance on the camera to custom. Take a photo of an 18% grey card or white dive slate underwater before the start of every SfM survey. Do this every time a new site is started.

NOTE: The photo will allow for color correction and will also help to separate the downloaded images from different sites when conducting multiple surveys on the same day.

2. Detailed methods

2.1. Site selection

2.1.1. Select a site that has enough room to swim the entirety of the spiral pattern (~113 m^2 in this example). In addition to the area being surveyed, incorporate a small buffer area to ensure that the entire survey area is sufficiently photographed to yield high-quality data.

2.1.2. Consider the ability and equipment of the two-person team; survey shallow sites ($\leq$~2 m) on snorkel and deeper sites by SCUBA.

2.2. If planning to repeatedly survey the site regularly, mark the center point, where the spool rig will be placed, with a tag or a permanent structure (e.g., rebar or cinder block). At the very least, take a global positioning system coordinate so that the site can be relocated with assistance from a printout of the orthomosaic.

NOTE: Permanent underwater structures typically require a permit.

2.3. Prepare the site.

2.3.1. Set the spool in the middle of the site.

2.3.2. Set out calibration tiles and record their depths. Place calibration tiles face-up, ~2 m away from the center.

NOTE: In this example, 3 calibration tiles were placed in a triangle around the center of the site. Calibration tiles should be appropriately weighted and positioned to ensure minimal movement during the collection of the photos.

2.4. Instruct the swimmer to swim with the camera while the assistant tends the spool.

2.4.1. Instruct the assistant to set the pole and the attached spool upright in the center of the selected site and hold the spool rig upright and stationary.

2.4.2. Ensure that the swimmer attaches the side of the camera closest to the spool to the line and holds the camera facing straight down ~1 m from the benthos.

NOTE: If the swimmer must tilt the camera, try to make sure that it is tilted slightly forward rather than backwards to avoid collecting images in the swimmer's shadow. Tilting the camera slightly forward for both the outward spiral and the return spiral may also capture better angles of the benthos and produce better models, especially when there are overhangs and holes.

2.4.3. Once the camera is properly positioned, instruct the swimmer to begin taking continuous images of the benthos while swimming forward and maintaining tension on the line.

2.4.5. Ensure that the swimmer continues to swim in a spiral at a consistent speed while taking photographs until the line is completely unwound from the spool.

NOTE: The swimmer should try to stay a constant distance of ~ 1 m above the benthos and swim the spiral at a moderate pace to ensure sufficient overlap between images. When in doubt, slower is better.

2.4.6. In highly rugose environments (e.g., coral reefs), include a third worker (second assistant) who can prevent line entanglement by hovering above the center of the line and gently lifting it over obstacles.

2.4.7. When the line is completely unspooled, instruct the swimmer to reverse directions, reattaching the camera if necessary, and swim the camera in the opposite direction to begin re-winding the line back onto the spool while taking pictures.

NOTE: Swimming the reverse spiral is not absolutely necessary, but will typically produce better models.

2.4.8. If a single spiral method is desirable to save time, instruct the swimmer to detach the line from the camera and skip to step 2.4.12 while the assistant winds the line and removes the spool rig from the site.

2.4.9. As soon as the swimmer begins to swim in the opposite direction, instruct the assistant to rotate the spool to wind the line in ½ of a turn (180°) against the new swimming direction to ensures that the swimmer's return path is offset from the original path to yield greater photo coverage of the site.

2.4.10. Ensure that the swimmer continues to take pictures and swim the reverse spiral until the line is almost completely rewound around the spool.

2.4.11. When the swimmer's and assistant's spacing prevents further progress, instruct the swimmer to stop taking pictures so that the camera can be detached from the line, and the assistant can remove the spool rig from the center of the site.

2.4.12. Once the spool is removed from the site, instruct the swimmer to image the center of the site by holding the camera facing straight down and moving the camera in a small spiral pattern over the center of the site.

3. Clean up the site.

3.1. Pick up calibration tiles and any other equipment before departing the site.

NOTE: Never leave trash or equipment at a site. Always leave a site cleaner than you found it.

Representative Results:

In this example, Reef Site 2--_7 located on Patch Reef 13 in Kāneʻohe Bay, Oʻahu, Hawaiʻi, was imaged, and 3,125 JPEG photos from the DSLR and 3,125 JPEG frame captures from the action camera video (**Table 1**) were used as input to create the orthomosaics and 3D models. The general workflow consisted of 5 stages: 1) alignment of photos to generate the sparse point cloud, 2) scaling the sparse point cloud and optimizing cameras, 3) building the dense point cloud (depth maps were also generated during this stage), 4) building the digital elevation model (DEM) and orthomosaic, and 5) generating the 3D model and texture. Note that stages 4 and 5 do not necessarily need to be done in that order, but they must be performed after processing the dense point cloud and depth maps. Georeferencing the models should occur prior to generating the orthomosaic and DEM. The settings used for these stages and processing details are outlined in **Table 2** and **Table of Materials**, respectively.

For more detailed methods of how to generate 3D models and orthomosaics see the **Supplementary Material** and Suka et al.[23]. Processing time was shorter for the action camera-derived model for every step including sparse point cloud generation, dense point cloud generation, mesh model rendering, and textured model rendering. This led to a significantly faster overall processing time for the action camera model (6 h 39 min) than the DSLR model (9 h 14 min). The exact time for model processing will vary with computational power and specific hardware configurations.

The model generated using images from the DSLR camera contained 2,848,358 sparse cloud points and 787,450,347 dense cloud points while the model generated from the action camera images contained only 2,630,543 sparse cloud points and 225,835,648 dense cloud points. This led to the DSLR models having more than 2x the resolution than the action camera models with orthomosaics resolutions of 0.442 and 0.208 mm/pixel for the DSLR- and action camera-derived models, respectively (**Table 1**). Despite the better resolution of the DSLR model relative to the action camera model, both methods were able to produce high-quality models with little difference in visual representation when the ~113 m^2 reef area was represented as a 20 cm^2 digital elevation model (**Figure 2 top panels**) or 2D orthomosaic projection (**Figure 2 middle panels**).

Figure and Table Legends:

Figure 1: Structure-from-Motion photogrammetry. (**A**) Example of a spool rig for controlling swimmer distance with an attached handle and pole for precise positioning and handling. (**B**) Calibration tiles. (**C**) A schematic of the swim path with relative positions of the swimmer (green) and the assistant (orange).

Figure 2: Visual comparison of digital elevation models and orthomosaics. Digital elevation models (top) and orthomosaics (middle) constructed from DSLR (left) and action camera (right) images. The bottom panel is a zoom of the areas in the white boxes in the orthomosaics. The heatmap scales in the top panel represent distance from the surface of the water in meters (m).

Table 1: Detailed information about setup cost, photos used to construct the models, photogrammetry metrics, and processing time. Processing was done using the same settings for both models. Note that processing time does not include time for various steps such as photo editing, extracting images from video, re-aligning photos, and editing and scaling the models.

Table 2: Detailed information on collected images and photogrammetric processing. Processing was done using the same settings for both models.

Discussion:

This study demonstrates that both the DSLR camera and the action camera produce models with better than 0.5 mm/pixel resolution in less than 10 h of processing time on a standard desktop computer. The major tradeoff between the DSLR and action camera, aside from cost, is finer resolution versus faster processing time, respectively. However, the reported processing times only include the computational processing. Thus, although the computational time is less for the action camera, there is a significant amount of time (10–20 min) invested in image extraction from the videos that is not required with the DSLR. An alternative is to use the action camera in continuous shooting mode to avoid image extraction. Continuous shooting mode was not used in this example, as the action camera can only shoot at 2 fps, which requires a significantly slower swim-rate to collect enough images to build a complete model. In this regard, there is a tradeoff between longer time in the field using the continuous shooting mode versus longer time on the computer, extracting images, when using video mode.

Advantages of the action camera include affordability and ease of transport and operation underwater. The main advantage of the DSLR is that it produces higher resolution images; hence, DSLR cameras are recommended over action cameras when the former is not cost-prohibitive. The kinds of questions a study seeks to address will also be important in determining the method used. For instance, an action camera might be preferable in environments that are relatively homogenous (e.g., seagrass beds, dead coral/rubble habitats), or where only broad community metrics (such as abundance, diversity) are being assessed over large spatial scales. However, a DSLR camera might be deployed in cases where tracking fine-scale changes in individual organisms or substrates is of interest.

As this is a field method, the model outputs will depend on various environmental factors such as lighting, water clarity, surface conditions, amount of surge, and movement of fish or non-stationary benthic structures (e.g., sea grass). Although there are no absolute thresholds of when it is appropriate to use this method, slightly overcast days with high water clarity, calm surface conditions, and little surge typically produce the best models. Moreover, there is a limit to the minimum depth required for these methods. These methods do not work well under conditions where there is less than 0.5 m of water because of the low overlap between photos and fewer distinguishing features per photo. However, this does highlight another advantage of the action camera, i.e., they are smaller and thus are easier for use at shallower depths. Furthermore, a smaller diameter spool and higher frame rate (or wider-angle lens) can improve image overlap in very shallow conditions[9].

Many other data types can be integrated with this approach. For example, orthomosaics have been used to show the spatial density of molecular data (e.g., genes and metabolites) on corals[24] and humans[25] using the open source software '*ili*'[26]. The same platform could also be used to map the spatial densities of animals, microorganisms, viruses, and/or chemicals in the environment. Other examples have used SfM for annotating benthic species spatially onto orthomosaics using geographic information system software[10]. Furthermore, the 3D models generated by SfM can be used to estimate habitat characteristics such as rugosity and fractal dimension. Indeed, the methods outlined here were recently used to derive a new geometric theory for habitat surfaces[10]. Finally, orthomosaics are being used as input surfaces for spatially explicit computational models, allowing for dynamical simulations to be overlaid on the model's 3D surface. Being able to easily generate large images and 3D representations of benthic habitats has allowed marine scientists to address hitherto unimagined questions[3].

Overall, here is a detailed protocol for conducting underwater SfM photogrammetry with either DSLR cameras or more cost-effective action cameras. These methods can be used by scientists for a broad range purposes, from extracting data about benthic ecosystems to developing 3D input surfaces for *in silico* simulations. However, these protocols can also be used by non-scientists as part of citizen science efforts to gather valuable information on patterns of biodiversity, habitat complexity, community structure, and other ecological metrics.

Acknowledgments:
We thank the Paul G. Allen Family Foundation for funding this research and are grateful to Ruth Gates for the inspiration to use technology to help conserve reefs. We also thank NOAA and other collaborators for thoughtful discussion concerning these methods. Lastly, we thank Catie Foley and Patrick Nichols for providing the drone and underwater video of these methods.

Disclosure:
The authors have no competing financial interests or other conflicts of interest.

References:
1. Levy, J., Hunter, C., Lukacazyk, T., Franklin, E. C. Assessing the spatial distribution of coral bleaching using small unmanned aerial systems. *Coral Reefs*. **37** (2), 373–387 (2018).
2. Muller-Karger, F. E. et al. Satellite sensor requirements for monitoring essential biodiversity variables of coastal ecosystems. *Ecological Applications*. **28** (3), 749–760 (2018).
3. Dornelas, M. et al. Towards a macroscope: Leveraging technology to transform the breadth, scale and resolution of macroecological data. *Global Ecology and Biogeography*. **28**, 1937–1948 (2019).
4. Burns, J. H. R., Delparte, D., Gates, R. D., Takabayashi, M. Integrating structure-from-motion photogrammetry with geospatial software as a novel technique for quantifying 3D ecological characteristics of coral reefs. *PeerJ*. **2015** (3), e1077 (2015).
5. House, J. E. et al. Moving to 3D: Relationships between coral planar area, surface area and volume. *PeerJ*. **2018** (6), e4280 (2018).
6. Carlot, J. et al. Community composition predicts photogrammetry-based structural complexity on coral reefs. *Coral Reefs*. 1–9, doi: 10.1007/s00338-020-01916-8 (2020).
7. Young, G. C., Dey, S., Rogers, A. D., Exton, D. Cost and time-effective method for multi-scale measures of rugosity, fractal dimension, and vector dispersion from coral reef 3D models. *PloS ONE*. **12** (4), e0175341 (2017).
8. Wilson, S. K., Robinson, J. P. W., Chong-Seng, K., Robinson, J., Graham, N. A. J. Boom and bust of keystone structure on coral reefs. *Coral Reefs*. **38** (4), 625–635 (2019).
9. Pizarro, O., Friedman, A., Bryson, M., Williams, S. B., Madin, J. A simple, fast, and repeatable survey method for underwater visual 3D benthic mapping and monitoring. *Ecology and Evolution*. **7** (6), 1770–1782 (2017).

10. Torres-Pulliza, D. et al. A geometric basis for surface habitat complexity and biodiversity. *Nature Ecology & Evolution.* **4**, 1495–1501 (2020).

11. Bayley, D., Mogg, A., Koldewey, H., Purvis, A. Capturing complexity: field-testing the use of 'structure from motion'derived virtual models to replicate standard measures of reef physical structure. *PeerJ*. **2019** (7), e6540 (2019).

12. Leon, J. X., Roelfsema, C. M., Saunders, M. I., Phinn, S. R. Measuring coral reef terrain roughness using 'Structure-from-Motion' close-range photogrammetry. *Geomorphology*. **242**, 21–28 (2015).

13. Burns, J. H. R. et al. Assessing the impact of acute disturbances on the structure and composition of a coral community using innovative 3D reconstruction techniques. *Methods in Oceanography*. **15**–**16**, 49–59 (2016).

14. Edwards, C. B. et al. Large-area imaging reveals biologically driven non-random spatial patterns of corals at a remote reef. *Coral Reefs*. **36**, 1291–1305 (2017).

15. Piazza, P. et al. Underwater photogrammetry in Antarctica: long-term observations in benthic ecosystems and legacy data rescue. *Polar Biology*. **42** (6), 1061–1079 (2019).

16. Bonney, R. et al. Citizen science: A developing tool for expanding science knowledge and scientific literacy. *BioScience*. **59** (11), 977–984 (2009).

17. Dickinson, J. L., Zuckerberg, B., Bonter, D. N. Citizen science as an ecological research tool: Challenges and benefits. *Annual Review of Ecology, Evolution, and Systematics*. **41** (1), 149–172, (2010).

18. Storlazzi, C. D., Dartnell, P., Hatcher, G., Gibbs, A. E. End of the chain? Rugosity and fine-scale bathymetry from existing underwater digital imagery using structure-from-motion (SfM) technology. *Coral Reefs*. **35** (3), 889–894 (2016).

19. Ventura, D., Lasinio, G. J., Belluscio, A., Ardizzone, G. A low-cost drone based application for identifying and mapping of coastal fish nursery grounds Feeding ecology View project Habitat use of juvenile Diplodus species View project. *Estuarine Coastal and Shelf Science*. **171**, 85–98, (2016).

20. Ventura, D., Bonifazi, A., Gravina, M. F., Belluscio, A., Ardizzone, G. Mapping and classification of ecologically sensitive marine habitats using unmanned aerial vehicle (UAV) imagery and object-based image analysis (OBIA). *Remote Sensing*. **10** (9), 1331 (2018).

21. George, E. E. et al. Relevance of coral geometry in the outcomes of the coral-algal benthic war. *bioRxiv*. doi: 10.1101/327031 (2018).

22. Anelli, M. et al. Towards new applications of underwater photogrammetry for investigating coral reef morphology and habitat complexity in the Myeik Archipelago, Myanmar. *Geocarto International*. **34** (5), 459–472 (2017).

23. Suka, R. et al. Processing photomosaic imagery of coral reefs using structure-from-motion standard operating procedures. U.S. Dept. of Commerce, NOAA Technical Memorandum NOAA-TM-NMFS-PIFSC-93, doi: 10.25923/h2q8-jv47 (2019).

24. Galtier d'Auriac, I. et al. Before platelets: the production of platelet-activating factor during growth and stress in a basal marine organism. *Proceedings of the Royal Society B: Biological Sciences*. **285** (1884), 20181307 (2018).

25. Bouslimani, A. et al. Molecular cartography of the human skin surface in 3D. *Proceedings of the National Academy of Sciences of the United States of America*. **112** (17), E2120–E2129 (2015).

26. Protsyuk, I. et al. 3D molecular cartography using LC-MS facilitated by Optimus and 'ili software. *Nature Protocols*. **13** (1), 134–154 (2018).

Figure 1

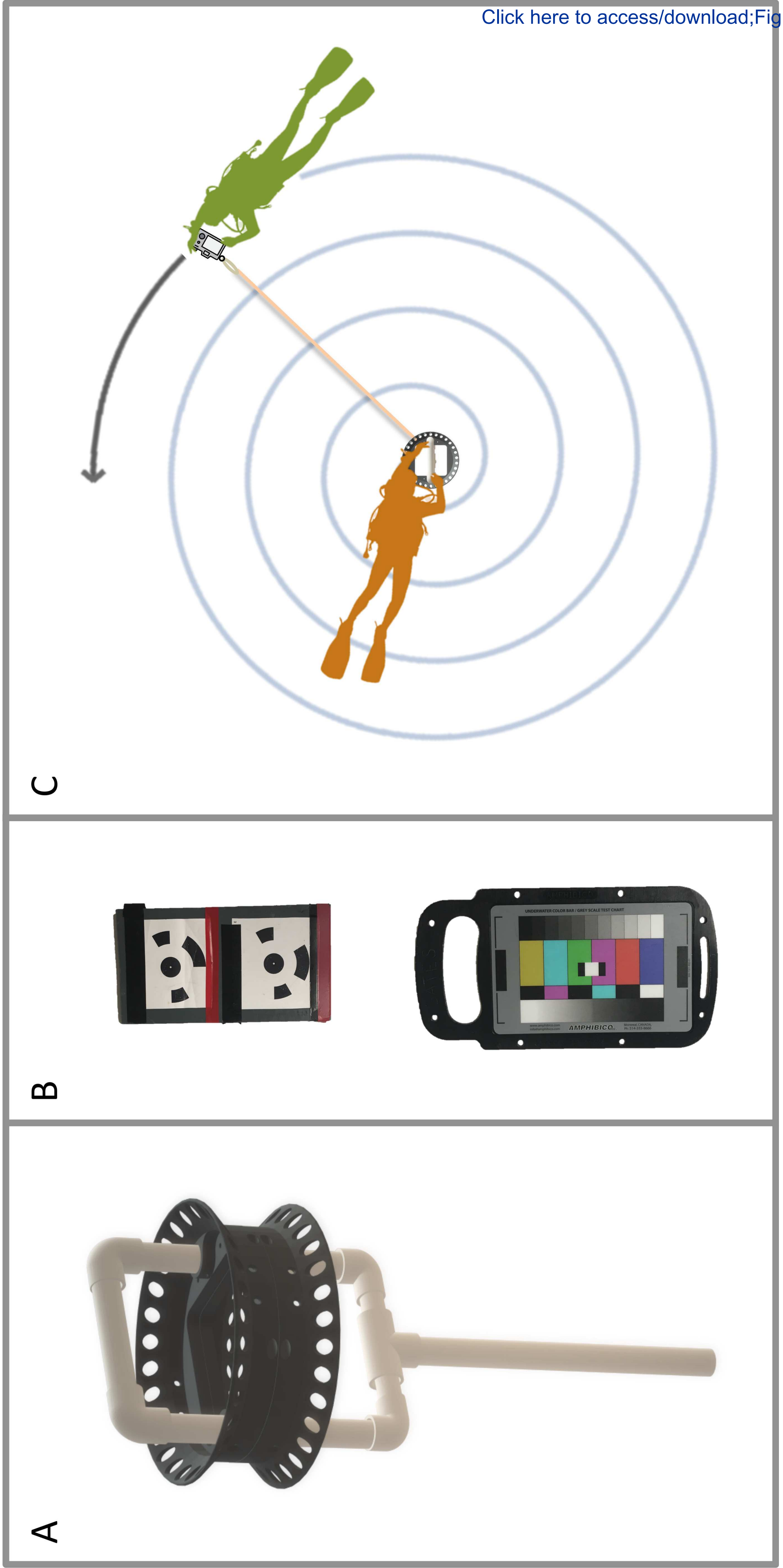

Figure 2

Click here to access/download;Figure;Final Figure 2.eps

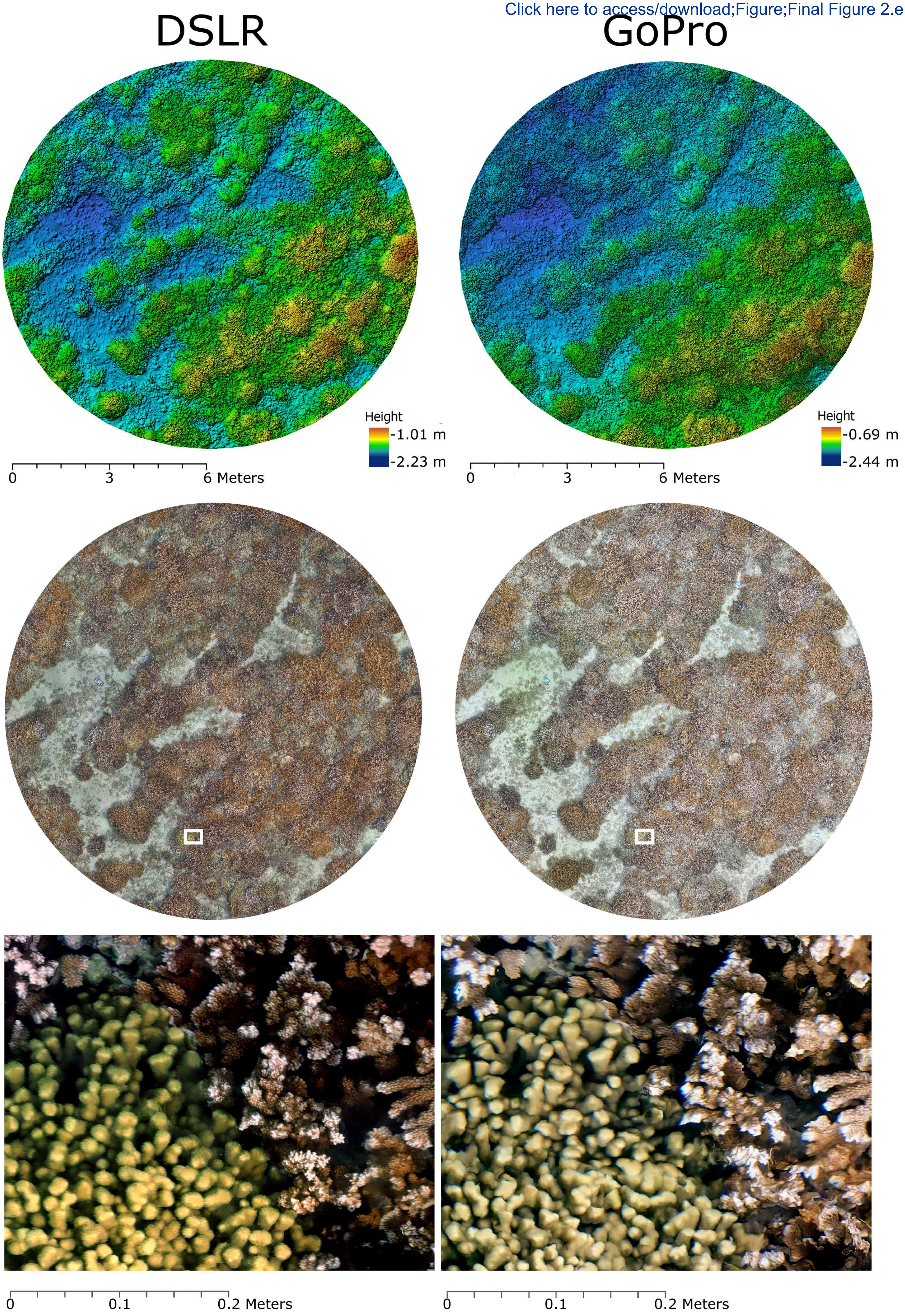

Table 1

Click here to access/download;Table;JoVE_Table_1.xlsx

	Canon EOS Rebel SL3	**GoPro Hero 7**
Cost		
Camera	~$600.00	~$220.00
Underwater housing	~$1,700.00	*NA*
Total Cost	**~$2,300.00**	**~$220.00**
Photos		
Photo file format	jpeg	jpeg
Photo resolution	24 Megapixels	12 Megapixels (from 4K video)
Aligned photos / total photos	3125 / 3125	3125 / 3125
Photogrammetry metrics		
Sparse cloud points	2,848,358	2,630,543
Dense cloud points	787,450,347	225,835,648
Faces (3D model)	11,919,451	3,834,651
Digital elevation model (DEM) resolution	0.831 mm/pixel	1.77 mm/pixel
Orthomosaic resolution	0.208 mm/pixel	0.442 mm/pixel
Processing times		
Sparse cloud generation	1 h 23 min	1 h27 min
Dense cloud generation	4 h	3 h 11 min
Mesh model rendering	3 h 32 min	1 h 49 min
Texture rendering	19 min	12 min
Total computer processing time	**9 h 14 min**	**6 h 39 min**

Table 2 Click here to access/download;Table;JoVE_Table_2.xlsx

	Canon EOS Rebel SL3	GoPro Hero 7
Images		
Average file size	~ 8.3 MB	~ 4.7 MB
Photo acquisition	Continuous mode	Extracted from 4K video
Color correction	Manual	Manual
Lens correction	No	Yes
Photogrammetry Process Settings		
Sparse cloud generation	Accuracy: High Key Point: 40,000 Tie Point: 4,000 Generic Preselection: Yes	Accuracy: High Key Point: 40,000 Tie Point: 4,000 Generic Preselection: Yes
Dense cloud generation	Medium Quality	Medium Quality
3D mesh model generation		
Source data:	Depth Maps	Depth Maps
Quality:	Medium	Medium
Face count:	Low	Low
Interpolation:	Enabled	Enabled
Calculate vertex colors:	Yes	Yes
3D texture generation		
Texture type:		Diffuse Map
Source data:	Diffuse Map Images	Images
Mapping mode:	Generic	Generic
Blending mode:	Mosaic	Mosaic
Texture size/count:	4096 / 1	4096 / 1
Digital elevation model (DEM)	From Dense cloud	From Dense Cloud
Orthomosaic	From DEM	From DEM

 Click here to access/download;Table of Materials;Revised_JoVE_Materials_Roach.xls

Name of Material/ Equipment	Company	Catalog Number	Comments/Description
Action camera (GoPro Hero7 Black)	GoPro		Could be any waterproof action camera
Adobe Lightroom	Adobe		Color correction
Calibration tiles (flat PVC board cut to size for Agisoft targets. Attach a dive weight underneath if expecting waves)			Any negatively buoyant object of known size and color. We recomme
DSLR camera (Cannon EOS Rebel SL3)	Cannon	3453C002AA	Could be any DSLR camera in a underwater housing
Line (plastic clothes line filament)			Any negatively buoyant line that is strong enough to withstand field u
Micro SDXC memory card (for GoPro)			
Oceanic Veo 2.0	Oceanic		Digital depth gauge
SDXC memory card (for DSLR)			Any SDXC memmory card should work, so long as there is enough spa
Spool (2 inch-long section of 8 inch diameter PVC pipe which was attached to a 3 feet section of 1 inch PVC pipe to form the stem			Any negatively buoyant, round object of the desired diameter
Underwater camera housing for DSLR (Ikelite 200DLM/C Underwater TTL Housing)	Ikelite	6970.09	Should be the specific water housing for the DSLR make and model
Windows 10 desktop computer with an Intel i9-9900K 8-core CPU, two Nvidia GeForceRTX 2070 SUPER GPUs, and 128 GB of RAM.			Processing

nd using the scale marker templates available from Agisoft Metashape software (v.1.6.0).

ıse

ce to hold all the pictures necessay to build the model

October 31, 2020

Dear JoVE Editors,

We are writing to re-submit our manuscript titled, *A field primer for monitoring benthic ecosystems using structure-from-motion photogrammetry* (MS tracking #: 61815) for publication in *JoVE*.

We have addressed the concerns of the editor and reviewers, and we appreciate your consideration of our revised manuscript. We would like to thank you and the reviewers for your comments. The critiques and suggestions have contributed to a more well-written paper, and a more accurate and understandable presentation of our method.

To address the reviewer's comments, we have made substantial improvements to the main text, the figures, and the Supplementary Material. In particular, we have provided formatting changes, as requested by the editor, and we have added to the discussion to more thoroughly address limitations and future applications of the technique.

Overall, we believe we have addressed the reviewers' comments and made the necessary changes, and that the manuscript is now suitable for publication in *JoVE*. Please find below a detailed response to the editor's and reviewers' comments. All responses are in bold.

Sincerely and on behalf of my coauthors,
Dr. Ty NF Roach

Dear Dr. Roach,

Your manuscript, JoVE61815 "A field primer for monitoring benthic ecosystems using structure-from-motion photogrammetry," has been editorially and peer reviewed, and the following comments need to be addressed. Note that editorial comments address both requirements for video production and formatting of the article for publication. Please track the changes within the manuscript to identify all of the edits.

After revising and uploading your submission, please also upload a separate rebuttal document that addresses each of the editorial and peer review comments individually.

Your revision is due by **Aug 31, 2020**.

To submit a revision, go to the JoVE submission site and log in as an author. You will find your submission under the heading "Submission Needing Revision". Please note that the corresponding author in Editorial Manager refers to the point of contact during the review and production of the video article.

Best,

Vineeta Bajaj, Ph.D.
Review Editor

Editorial comments:

Changes to be made by the Author(s):

1. Please take this opportunity to thoroughly proofread the manuscript to ensure that there are no spelling or grammar issues. The JoVE editor will not copy-edit your manuscript and any errors in the submitted revision may be present in the published version.

We have carefully reviewed the manuscript for all spelling and grammar issues.

2. Please format the manuscript as: paragraph Indentation: 0 for both left and right and special: none, Line spacings: single. Please include a single line space between each step, substep and note in the protocol section. Please use Calibri 12 points.

We have formatted the manuscript as requested.

3. Please remove the running title from the manuscript.

The running title has been removed.

4. Please make the title concise.
The title is as concise as it can be to still communicate the necessary information.

5. Please provide an email address for each author.
Emails have been provided for each author in the editorial manager, and have also been provided below:
Ty Roach: smokinroachjr@gmail.com
Shreya Yadav: shreyay@hawaii.edu
Carlo Caruso: carloxcaruso@gmail.com
Jenna Dilworth: jenna.dilworth@rsmas.miami.edu
Catherine M. Foley: cmrfoley@gmail.com
Joshua R. Hancock: hancock9@hawaii.edu
Joel Huckeba: joel_huckeba@sbcglobal.net
Ariana S. Huffmyer: ashuff@hawaii.edu
Kira Hughes: kirawa@hawaii.edu
Valerie A. Kahkejian: vkahkejian@gmail.com
Elizabeth M.P. Madin: emadin@hawaii.edu
Shayle B. Matsuda: shayle@hawaii.edu
Michael McWilliam: mjmcwilliam@outlook.com
Spencer Miller: smiller11@my.hpu.edu
Erika P. Santoro: erikapsantoro@gmail.com
Mariana Rocha de Souza: mrds@hawaii.edu
Damaris Torres-Pulliza: dama.reefs@gmail.com
Crawford Drury: druryc@hawaii.edu
Joshua S. Madin: jmadin@hawaii.edu

6. Please ensure that the long Abstract is within 150-300-word limit and clearly states the goal of the protocol.
The abstract is only 152 words.

7. The Protocol should contain only action items that direct the reader to do something. Please move the overview section to the introduction instead.
The methods overview has been moved to the last part of the introduction.

8. Please ensure that all text in the protocol section is written in the imperative tense as if telling someone how to do the technique (e.g., "Do this," "Ensure that," etc.). The actions should be described in the imperative tense in complete sentences wherever possible. Avoid usage of phrases such as "could be," "should be," and "would be" throughout the Protocol. Any text that cannot be written in the imperative tense may be added as a "Note." However, notes should be concise and used sparingly.

The protocol section has been written in the imperative mood when possible. However, due to the multi-person nature of this protocol there are some sentences that could not be formulated in imperative.

9. Please add more details to your protocol steps. Please ensure you answer the "how" question, i.e., how is the step performed?
We have added more detail to the protocol when necessary.

10. We cannot have nonnumbered steps or headings in the protocol section e.g., Materials and Detailed method. Protocol should contain all numbered steps and subheading and should follow continuous numbering pattern.
All headings in the protocols section have been numbered.

11. Please use complete sentences throughout to explain how you perform your step with all specific details associated with it.
All steps in the protocol have been written in complete sentences where possible.

12. Please ensure that individual steps of the protocol should only contain 2-3 actions sentences per step.
All steps in the protocol contain 3 or fewer action items.

13. How was DSLR camera set up?
Details of how to set up the DSLR camera are provided in section 1.1.2 of the protocol.

14. Lines 227 -229: How is this dependent on the images obtained from different camera types. Need more details on these.
This has been addressed in the discussion.

15. Please include how the images were used to generate 3D models and orthomosaics.
Images were uploaded into a proprietary structure-from-motion photogrammetric software where point cloud, dense cloud, mesh model, digital elevation model, and orthomosaic were produced. This has all been included in the main text of the manuscript.

16. JoVE cannot publish manuscripts containing commercial language. Please remove all commercial language from your manuscript and use generic terms instead. All commercial products should be sufficiently referenced in the Table of Materials and Reagents.
For example: Gopro, Agisoft Metashape Pro (v. 1.6.0; Agisoft, LLC.), etc.
All references to commercial products have been removed.

17. There is a 10-page limit for the Protocol, but there is a 2.75-page limit for filmable content. Please highlight 2.75 pages or less of the Protocol (including headings and spacing) that identifies the essential steps of the protocol for the video, i.e., the steps that should be visualized to tell the most cohesive story of the Protocol. If highlighting underwater steps, please include the footages along with the revision.

As the entire protocol is less than 2.75 pages, we would like to include all steps as essential steps for the video.

18. Please ensure the results are described in the context of the presented technique. e.g., how do these results show the technique, suggestions about how to analyze the outcome, etc. The paragraph text should refer to all of the figures. Data from both successful and sub-optimal experiments can be included.

All figures are referenced in the results section, and results are described in the context of the protocol we present.

19. Please include all the Figure Legends together at the end of the Representative Results in the manuscript text. Each Figure Legend should include a title and a short description of the data presented in the Figure and relevant symbols. The Discussion of the Figures should be placed in the Representative Results.

Figure captions were placed sequentially at the end of the *Representative Results* Section in the main text of the manuscript.

20. Please obtain explicit copyright permission to reuse any figures from a previous publication. Explicit permission can be expressed in the form of a letter from the editor or a link to the editorial policy that allows re-prints. Please upload this information as a .doc or .docx file to your Editorial Manager account. The Figure must be cited appropriately in the Figure Legend, i.e. "This figure has been modified from [citation]."

No figures were reused from any previous publications.

21. As we are a methods journal, please ensure that the Discussion explicitly covers the following in detail in 3-6 paragraphs with citations:
a) Critical steps within the protocol
b) Any modifications and troubleshooting of the technique
c) Any limitations of the technique
d) The significance with respect to existing methods
e) Any future applications of the technique

All of the above discussion points have been explicitly covered in the discussion section of the manuscript.

22. Please include a Disclosures section, providing information regarding the authors' competing financial interests or other conflicts of interest. If authors have no competing financial interests, then a statement indicating no competing financial interests must be included.
A Disclosure section has been included in the manuscript just before the references. This section now states: "*The authors have no competing financial interests or other conflicts of interest.*"

23. Please remove the embedded figure(s) from the manuscript. All figures should be uploaded separately to your Editorial Manager account. Each figure must be accompanied by a title and a description after the Representative Results of the manuscript text.
All embedded figures have been removed from the manuscript and have been uploaded separately in the Editorial Manager.

24. Please remove the embedded Table from the manuscript. All tables should be uploaded separately to your Editorial Manager account in the form of an .xls or .xlsx file. Each table must be accompanied by a title and a description after the Representative Results of the manuscript text.
All tables have been de-embedded from the manuscript and each table has been individually uploaded in the Editorial Manager as a .xlsx file. The corresponding table captions have been added to the end of the Representative Results Section just after the Figure Captions.

25. Please sort the materials table in alphabetical order.
The material table has been sorted in alphabetical order and re-uploaded.

Reviewers' comments:

Reviewer #1:

Although the comparison between DSLR and GoPro cameras leads us to an obvious conclusion (DSLR cameras are better than action cameras when there are no funding limitations), this manuscript provided by T.N.F. Roach and contributors gives us a sharply enough protocol for conducting underwater Structure-from-Motion (SfM) surveys. SfM photogrammetry is a relatively low-cost, simple, non-invasive, and repeatable method that should be expanded and used or at least known by most marine benthic ecologists. The introduction is really well set, I've read it with a great interest !
Thank you for the positive feedback. We are happy to hear that you recommend the manuscript for publication.

Therefore, according to the work done here, the effort to be concise and clear, I recommend this manuscript with only one minor edit from my side (or maybe didn't I understand it well ?). concerning the 3rd paragraph from the results: "This led to the DSLR models having more than 2-times the resolution than the GoPro models (Table 1).", Would you not say 3-times, rather than twice? Overall, it is a great ms, well done!

This is a valid point, given the data we present in the text of the Results section. However despite the dense cloud from the DSLR model being ~ 3-times denser than the action camera model, the orthomosaic resolution is only ~2-times better (.442 mm/pixel for DSLR vs 0.208 mm/pixel for action camera). To make this more clear and understandable, we have added the actual orthomosaic resolution results to the main text in the Results Section. This section of the text now reads as follows:

> ***"The model generated using images from the DSLR camera contained 2,848,358 sparse cloud points and 787,450,347 dense cloud points while the model generated from the action camera images contained only 2,630,543 sparse cloud points and 225,835,648 dense cloud points. This led to the DSLR models having more than 2-times the resolution than the action camera models with orthomosaics resolutions of 0.442 and 0.208 mm/pixel for the DSLR and action camera derived models respectively."***

<u>Reviewer #2:</u>

Manuscript Summary:

I have read the paper entitled: "A field primer for monitoring benthic ecosystems using structure-from-motion photogrammetry" by Roach et al.

I found this methodological work very interesting because it provides a very innovative sampling techniques based on Structure from Motion (SfM) photogrammetry. The authors well describe the use of two camera models (a DSLR and an action cam) coupled with a spool ring to make underwater spiral paths the acquisition of overlapping images. Both methods allowed the reconstruction of a circular area (approximately 120 m2) of a coral reef in O'ahu (Hawaii) Island.

Thank you for your positive feedback.

I appreciate the clarity of the manuscript but I suppose that it should be improved by detailing much more the introduction and the conclusion sections, by explicating the innovation linked to this methodology and because it is more effective that classical visual estimation carried out by SCUBA divers. The manuscript could be suitable for publication after major revision.

We have added to both the introduction, conclusion, and discussion sections to better detail the innovations linked to this method, and to better contextualize its uses.

Major Concerns:
Specific comments:
Some important references on SfM photogrammetry applications are missing and should be added to better motivate some sentences in the introduction section such as:
"For examples of studies that have applied variations of this method to study underwater ecological communities…" or "The models generated from SfM can be used to collect data on the structural complexity (e.g., rugosity, dimensionality).."
Ventura, D., Bonifazi, A., Gravina, M. F., Belluscio, A., & Ardizzone, G. (2018). Mapping and classification of ecologically sensitive marine habitats using unmanned aerial vehicle (UAV) imagery and object-based image analysis (OBIA). Remote Sensing, 10(9), 1331.
Ventura, D., Bruno, M., Lasinio, G. J., Belluscio, A., & Ardizzone, G. (2016). A low-cost drone based application for identifying and mapping of coastal fish nursery grounds. Estuarine, Coastal and Shelf Science, 171, 85-98.
Casella, E., Collin, A., Harris, D., Ferse, S., Bejarano, S., Parravicini, V., ... & Rovere, A. (2017). Mapping coral reefs using consumer-grade drones and structure from motion photogrammetry techniques. Coral Reefs, 36(1), 269-275.
Leon, J. X., Roelfsema, C. M., Saunders, M. I., & Phinn, S. R. (2015). Measuring coral reef terrain roughness using 'Structure-from-Motion'close-range photogrammetry. Geomorphology, 242, 21-28.
Raoult, V., Reid-Anderson, S., Ferri, A., & Williamson, J. E. (2017). How reliable is Structure from Motion (SfM) over time and between observers? A case study using coral reef bommies. Remote Sensing, 9(7), 740.
Bayley, D. T., Mogg, A. O., Koldewey, H., & Purvis, A. (2019). Capturing complexity: field-testing the use of 'structure from motion' derived virtual models to replicate standard measures of reef physical structure. PeerJ, 7, e6540.
Thank you for pointing out these references. They have been cited here and added to the references list.

Moreover, many underwater surveys in tropical environments have been carried out to investigate reefs topology with the "chain and tape" method (See also: Young, G. C., Dey, S., Rogers, A. D., & Exton, D. (2017). Cost and time-effective method for multi-scale measures of rugosity, fractal dimension, and vector dispersion from coral reef 3D models. PloS one, 12(4), e0175341.). Thus, the authors could shed more light on the DSM generated to make a comparison with this more traditional method.

Protocol
1. Camera.

DSLR camera settings are referred to still images but for the Gorpro 7 authors used frames form 4K video sequences. Why they have not used time lapse mode to acquire still images with the action cam and why the raw format was not chosen for any of the cameras? Please clarify.
Time lapse setting was not used because there needed to be the same number of images from both cameras to provide a valid comparison. Time lapse on the GoPro produces only 2 frames per second while the Canon DLSR take ~4 frames per second. Using the 4K video setting and extracting frames (same resolution as time lapse photos) allowed to subsample an equal number of GoPro images as the DSLR images while not sacrificing the total number of images, or having to swim extremely slowly.

2. Spool rig
*If the author used a line of 6 m which implied a 12 m diameter, why the surveyed area is about 120 m2. With a radius of 6 m the area investigated should be: A = π * r 2 =3.14 * 36 = 113 m2.*
Thank you for catching this. You are correct that the area of the mosaic is ~113 m^2. This has been corrected throughout the manuscript.

3. Calibration tiles
The depth (z factor) of each coded target was estimated by a digital depth gauge (please add the model and its accuracy).
We have added the model (Oceanic Veo 2.0) in parentheses after referencing the digital depth gauge.

However, how the x and y were estimated? A GPS system is used from the surface? Please clarify.
***You are correct,* the x and y coordinates were estimated using a handheld GPS system from the surface.**

4. Color correction
Color correction was done with a third part software, different from Metashape, I suppose so add some information on parameters used.
Color correction was done manually in Adobe Lightroom. This has been added to the main text of the manuscript and has been explicitly described in section 1.4.1.

Results:
The positional accuracy of the model in not assessed
I suppose that the model generated in a DSM and not DEM
The photomosaic spatial resolution is not indicated.
The photomosaic resolution has been reported in the main text and in Table 1.

Processing time are useless if the hardware used is not specified. RAM, CPU and graphic Card are key elements that influence the total processing time.
The following sentence has been added to the results: "*Processing was performed on a Windows 10 desktop computer with an Intel i9-9900K 8-core CPU, two Nvidia GeForceRTX 2070 SUPER GPUs, and 128 GB of RAM.*"

Minor Concerns:
Figures:
The overall quality of figures should be improved.
The final version of the figures has been uploaded according to the journal's standards

Add the length radius on Fig. 1C
As we are aiming to make this as generalizable as possible, we have chosen not to add the length radius of our example as this length can be varied depending on the needs of the SfM team (this has been added to the protocols section).

Add a figure showing the study site and a table with habitat characteristics such as dominant species of corals and depth.
As figure 1 shows the details of the benthic structure of the study site, and the specifics of the site are not relevant to the manuscript (it is a methods paper), we have decided to leave out any additional figures concerning the specific site surveyed as our example.

Reviewer #3:
Manuscript Summary:
The authors describe the protocol for creating underwater "cookie" structure from motion models in coral reef environments. In parallel, the authors compare the results obtained from this method using two different types of camera, a DSLR with a mechanical shutter and an action camera with a rolling digital shutter. The manuscript is detailed and should allow researchers to replicate this approach elsewhere. My comments are relatively minor and should be easy to address.
Thank you for your positive feedback!

Major Concerns:
While the authors do a great job presenting the approach, I feel the language used oversells where the approach can be used. In many benthic habitats, notably temperate reefs where algae are the dominant structure, structure from motion does not work effectively because movement of the soft structures prevents image alignment. I think it would be more transparent that the authors rephrase 'benthic habitats' to 'coral reefs' as this is really where this method would work

best. This is really made evident by the citations referred to that exclusively use this method in coral reef environments.

We have left in the term, "benthic habitats" to emphasize that this approach can be used in many aquatic ecosystems, not just coral reefs. However, we have added a short sentence, which clearly states that if there is significant movement due to non-static benthic organisms, then this method may not work due to improper photo alignment for the models.

In addition, it's not clear to the reader why the cookie method is presented over others. I would suggest highlighting it produces more reliable results than traditional transects or S- shaped patterns, even if it comes at a cost of smaller sampling areas and a little more setup.

We have emphasized the advantages of this method over others by adding the following sentences: "*There are several methods to gather images for underwater SfM. The method described here has the advantage of being more repeatable, and controlling the swimming path, which is especially advantageous in rough, or surgy conditions.*"

I'm not familiar with the journal requirements, but I think more justification for comparing DSLR to action camera may be necessary for the reader to understand the reason for this discussion. Presumably, researchers will want the highest quality model possible with the lowest GSD, so is the premise higher quality vs affordability? My other major concern with this is that there are many studies that have already determined action cameras were fine for these purposes (e.g. Young et al. 2017, Raoult et al. 2017), so it's not immediately clear what the reason for this is. Processing time and accuracy of photogrammetry is largely a function of processing power and image resolution, so it might be better to frame the question around sampling resolution rather than action camera and DSLR? Presumably the lower resolution GoPro images could be processed at higher accuracy in Metashape to obtain similar results to the DSLR with a similar processing time. In addition, processing time is mostly determined by the machine that is conducting the image alignment, so the authors need to present detailed information on the PC used (CPU, RAM, graphics card) to be able to get an idea of their likely processing times.

We have added the detailed information on the PC in the results section as requested. This sentence reads as follows: "All processing was performed on a Windows 10 desktop computer with an Intel i9-9900K 8-core CPU, two Nvidia GeForceRTX 2070 SUPER GPUs, and 128 GB of RAM."

To address concerns about the tradeoffs between the two cameras, we have added a substantial amount to the discussion section, which we hope will remedy these concerns.

Minor Concerns:

Running title: "methods" is inaccurate, it is a single method.

The running title has been removed at the request of the editor.

Paragraph 3 of the intro: "up to speed with sfm surveys" again, the authors here present the protocols of just one approach used for SfM studies on benthic environments. The authors need to not oversell their research, the method is robust enough by itself.

The wording has been changed here to reflect that only a single method for SfM surveys is being presented. This line now reads: "*...up-to-speed with a simple benthic SfM survey method…"*

1.2.2: Focal length of the lens is a critical aspect for this (since it largely determines overlap and relates to image distortion), the authors should highlight this here. Was the focal length at the minimum (wide angle) or maximum (linearl)? There are similar settings available on the GoPro that should also be presented. Some studies have found little difference between viewing angles, and I believe wider angles help get more overlap and more reliable alignment.

We have added the focal length in the protocols section (1.1.2.2.2).

I think surface conditions should be discussed. Very sunny days in shallow areas cause problems with image alignment because of the rapid changes in lighting/colour. I would be recommending this method for depths >= 1.5m to reduce this effect.

The effects of depth, lightning, and other environmental factors have been added to the discussion.

1st paragraph of representative results: specify hardware that includes CPUs with more cores, higher frequency, more RAM, and graphics cards with CUDA cores can significantly reduce processing time. An example "achievable" appropriate configuration would be relevant. May also be relevant to include continuous shooting speed of Rebel (5fps) vs the GoPro (2fps).

The relevant hardware specifications have been included in the results section, stating that, "*All processing was performed on a Windows 10 desktop computer with an Intel i9-9900K 8-core CPU, two Nvidia GeForceRTX 2070 SUPER GPUs, and 128 GB of RAM.*"

3rd paragraph of representative results: I disagree there is very little difference in the models. The DEMs suggest the northern-most sections are quite different between the two models, with the GoPro having much lower surfaces. The min/max heights vary by ~40cm, which is more than a typical acropora would grow in a few years. I'd suggest this warrants highlighting that users stick to one type of camera so that results are comparable between sampling events (evidence that this is appropriate in Bryson et al. 2017 and Raoult et al. 2017).

We have removed the adverb "very" and we have also detailed possible reasons for this discrepancy in depth. The main reason being that the GoPro's lens is further from the benthos, and the tide came in ~1.5 hrs between the time we sampled with the DSLR and when we came back with the GoPro. The combined effects of the tide and the camera's size yielded a GoPro model with a slightly greater depth.

May be relevant to highlight in methods that artificial lights/flash is usually not appropriate for SfM and users should avoid them.
This has been added to the methods (Section 1.2.2.11 reads: "*Note: Artificial lights and/or flashes should be turned off while collecting SfM images*").

2.3: Is there any reason users couldn't use a longer or shorter radius? Might be useful to state.
The sentence, "*However, the length of the line can be varied in order to capture images from a desired area of a given size.*" has been added to the end of section 2.3 in the methods.

3.5.1 May be better to state "at a constant height at least 1m above benthos", since the important aspect here is staying at a height sufficient to allow good image overlap.
Thank you for this suggestion. This change has been made. The sentence now reads: "*The swimmer should try to stay a constant distance of at least 1 m above the benthos*".

Specifying that users should incorporate a 'buffer' area around their desired shape (i.e. if radius = 3m, start at ~4m to make sure 3m data is good quality), commonly employed in drone surveying.
This is a good point. The following sentence has been added here: "*In addition to the area being surveyed, incorporate a small buffer area to ensure that the entire survey area is sufficiently photographed to yield high-quality data.*"

Discussing image number/MP is necessary in discussion. Larger image number/MP images reaches limits of models and typical PC processing, especially in Metashape which is very RAM-dependent, so users might prefer to have multiple smaller cookies rather than on larger one (since the larger one will be difficult to process).
This and other relevant discussion points have been added to the *Discussion* section of the manuscript.

Discussion paragraph 2: I'm not sure I agree with statements re action cameras. It might be more honest to say "If authors want the absolute highest resolution, then a DSLR with a higher resolution sensor (there are some with >40MP now) is optimal" but I don't think they can state a GoPro is always inferior. The higher resolution MP of the DSLR limit survey size to some degree, so a larger area could be surveyed with the GoPro without running into hardware issues for processing. It might be preferable to have more than one camera so sampling can occur simultaneously, in which case the action camera would be preferable. In addition, many groups use setups with multiple action cameras attached to a pole which allows better image alignment and covering larger areas more rapidly (even if swim speed needs to be slower). Like everything, there are trade-offs so I don't believe it's so clear-cut. I'm also not aware of any studies using SfM on seagrass, probably because getting image alignment will be near impossible since these are mobile structures, but if authors have an example they should include it here.

I would suggest not citing papers that are preprints. If the COVID crisis has taught us anything, it's that while preprints are fantastic for making research available and debatable, they are not really appropriate to build further research upon. There's plenty of material available that addresses the authors' points.
This reference has been removed.

<u>Reviewer #4:</u>
Manuscript Summary:
This manuscript explains the workflow for conducting underwater structure-from-motion (SfM) photogrammetry surveys of benthic habitats in a clear water site (Kāneʻohe Bay, Oʻahu, Hawaiʻi). The manuscript includes a comparison of two cameras (DSLR and GoPro) to assess differences in computational time and resolution. The manuscript provides a detailed protocol for conducting underwater SfM surveys and descriptive results of the tradeoffs between each collection method. The manuscript provides a repeatable protocol for scientists and non-scientists to gather valuable information on benthic habitats. With revisions, and potentially an addition of an additional scientific analysis, this manuscript with an accompanied video would assist others and contribute to the growing research in underwater photogrammetry.
Thank you for your positive feedback.

Major Concerns:
** Looking at the JoVE instructions for authors, I am unsure if the "Materials" and "Detailed Methods" are required in the Protocol section. The authors should clarify that both lists are allowed. Some of the text in the materials and detailed methods are repetitive. I would suggest removing actionable items in the materials list, and only listing them in the detailed methods.*
These sections have been edited to meet journal standards.

** The authors should consider adding text about the required clarity of the water and whether this method be conducted in turbid water bodies or tidal estuaries. Text incorporating information or answers to the following questions would improve the quality of this manuscript:*
o Is there a clarity threshold for effective underwater photogrammetry?
o Is there a certain time of season that provides the most clarity?
o Is there a particular time in a tidal cycle where surveys should be completed (e.g. slack tide)?
o What type of ambient lighting is recommended? (e.g. Do cloudy/overcast days effect underwater lighting?)
These are very good discussion points. As such we have added an additional paragraph to the discussion section, which reads as follows:

> **"*As this is a field method, the model outputs will depend on various environmental factors. These factor include, water clarity, surface conditions, amount***

> ***of surge, and movement of fish or non-stationary benthic structures (e.g., sea grass). Although there are no absolute thresholds of when it is 'OK' to use this method, we find that slightly overcast days with high water clarity, calm surface conditions, and little surge typically produce the best models. We also note that there is a limit to the minimum depth required for these methods. We find that under any conditions where there is less than 0.5 m of water, these methods do not work well. However, this does highlight another advantage of the action camera - they are smaller and thus are better for use at shallower depths.*"**

** I have some concerns over the robustness of the manuscript. I am unsure if the comparison of computational time and resolution from two cameras is of enough scientific value. Perhaps the authors could include a comparison of the DEMs to the actual in situ benthic heights. A comparison of a benthic complexity metric such as rugosity could also be included. If in situ data (chain and tape) can be collected, height or rugosity values for each DEM can be compared to determine best performing camera and processing parameters. Perhaps the analysis methods of Storlazzi et al. (2016) or Burns et al. (2015) can be used to add scientific value to this manuscript.*

As this is simply a methods paper, we have chosen not to collect additional data types, as we feel that this would only confuse the general reason for the paper.

Minor Concerns:

Abstract

Line 26: define DSLR

We have spelled out the term (Digital Single Lens Reflex) in the abstract.

Introduction

Line 37: Consider changing to "such as 3D laser scanning…."

We have changed this line accordingly.

Line 37: Consider changing to "satellite remote sensing" instead of "satellite imagery"

We have changed from "satellite imagery" to "satellite remote sensing" as you requested.

Line 40: Delete "some"

The word "some" has been deleted here.

Line 47: Already defined what SfM means in above paragraph, so only need SfM here.

As this term has only been defined in the abstract, we have left the parenthetical definition as this is the first mention of the term in the main body of the manuscript.

Line 54: Change to "valuable and objective"

This has been changed in the manuscript.

Line 53: These adjectives are repetitive to beginning first sentence in paragraph. Consider taking them out so it reads "Furthermore, this method can be used by…"
We have chosen to leave these adjectives in the manuscript, to further emphasize the benefits of this approach

Line 58: Consider changing to "This study provides a detailed…."
This line has been changed. It now reads: "*This work provides a detailed*…"

Lines 58-60. Delete "also". Consider adding the specific DSLR and action camera in parenthesis.
The word "also" has been removed here. The specific camera types cannot be added, as per journal policy.

Line 61: Consider changing "up-to-speed" to "informed with updated benthic SfM surveys…"
This line has been changed according to your request.

Protocol:
Line 66: Add information on required scuba diving or snorkeling. "The method described here requires a two-person snorkeling or scuba diving team."
The words, "Snorkel or SCUBA" have been added. This sentence now reads: "*The method described here requires a two-person snorkel or SCUBA team*."

Line 67: Consider adding recommended length of line in parenthesis
We have decided not to add a recommended length of line, as the length of the line can be varied depending on the needs of the survey team.

Line 67: Add "underwater" after "placed". How is it placed? Consider adding the role of the assistant here.
The word "underwater" has been added here.

Line 67: Add number or range of number of calibration tiles
This has been added to the detailed methods portion of the protocol

Materials:
Line 127: Consider adding a link to the spool 3D print file in this section or in supplementary materials.
We have added a link in the SI.

Line 134: Add "or wrist" after "camera".
We have chosen not to add the words “or wrist”, as in our example the line is always attached to a detachable clip, even when it is subsequently attached to the diver’s wrist.

Line 137: Change "but" to "however"
“But” has been changed to “however” here.

Line 141: Consider adding link to Agisoft scale marker templates either in this section or supplementary material
This has been included in the SI.

Line 162: Change "Note," to "NOTE:"
This has been changed here and throughout the manuscript.

Representative Results:
Line 215: Is there a figure to reference in the first sentence? If not then just say "For our example, we imaged six sites on a coral reef located in…."
There is not a figure to reference in this sentence. We have changed the sentence to better communicate that the reef site, 2_7, was a single site located on Reef 13 in Kaneohe Bay, HI.

Line 218: Remove "Briefly"
The word “briefly” has been removed here.

Line 220: Consider adding the word "point" each time you reference a "dense cloud" or sparse cloud".
We thank you for this suggestion and have added the word “point” here and elsewhere when referring to dense and sparse clouds.

Line 232-233. Consider calling the "cloud points" "keypoints" instead. This is common terminology in photogrammetry.
Thank you for this suggestion, but we have chosen to keep the phrase “cloud points” as this is a more generic term than “key points”.

Line 238: This sentence is unclear. How was the 120m2 reef are represented as a 20cm2 DEM or ortho projection?
The reef area was represented as both a 20 cm2 DEM and a 2D orthoprojection. We have added which figure panels are being referred to in this sentence to help clarify this point. This sentence now reads: “*Despite the better resolution of the DSLR model relative to the*

***action camera model, both methods were able to produce high-quality models with very little difference in visual representation when the 113 m2 reef area was represented as a 20 cm2 digital elevation model (Figure 2 top panels) or 2D orthomosaic projection (Figure 2 middle panels).*”**

Discussion and conclusion
Line 247: Include the length of GoPro video and how long it took to extract images so readers get a an understanding how long this tasks takes on top of processing time. (Does Agisoft have a video import function? I know Pix4D does. If it does, this could eliminate image extracting time).
The length of the GoPro video was approximately 14 minutes. We have added how much time this took in this part of the discussion (~15 min). In light of editorial review, we have not mentioned using proprietary software such as *Agisoft* or *Pix4D* in the discussion, as this is against journal policy.

Line 267: Missing a quotation after "ili". Add more information on this open source platform. Change "platform" to "software".
We have added the quotation after “ili” and changed the word “platform” to “software”.

Line 274: Use SfM abbreviation
This has been changed to the SfM abbreviation.

Figure 2. Make DEM legends the same scale. Is this height from surface of water? Clarify this in figure caption.
This scale is a measure of distance from the surface of the water (m). This has been clarified in the figure caption.

Table 1:
Line 359: include Agisoft version in caption.
- Not sure what "faces" of a 3D model are. You should include this information in the text.
- Consider adding cost of an Agisoft license or alternative photogrammetry software (e.g. Pix4D)
All references to Agisoft and specific softwares have had to be removed due to journal specifications.