

# Measurement of Micro-bathymetry with a GOPRO Underwater Stereo Camera Pair

Val E. Schmidt and Yuri Rzhanov  
Center for Coastal and Ocean Mapping  
University of New Hampshire  
Durham, NH

**Abstract**— A GO-PRO underwater stereo camera kit has been used to measure the 3D topography (bathymetry) of a patch of seafloor producing a point cloud with a spatial data density of 15 measurements per 3 mm grid square and an standard deviation of less than 1 cm. A GO-PRO camera is a fixed focus, 11 mega-pixel, still-frame (or 1080p high-definition video) camera, whose small form-factor and water-proof housing has made it popular with sports enthusiasts. A stereo camera kit is available providing a waterproof housing (to 61 m / 200 ft) for a pair of cameras. Measures of seafloor micro-bathymetry capable of resolving seafloor features less than 1 cm in amplitude were possible from the stereo reconstruction. Bathymetric measurements of this scale provide important ground-truth data and boundary condition information for modeling of larger scale processes whose details depend on small-scale variations. Examples include modeling of turbulent water layers, seafloor sediment transfer and acoustic backscatter from bathymetric echo sounders.

**Index Terms**—stereo imaging, seafloor bathymetry, acoustic backscatter

## I. INTRODUCTION

In many areas of oceanographic study measures of the seafloor on a very small scale (capable of resolving variations of just a few mm) are desired. For example, modeling of meso-scale turbulence across the seafloor requires a statistical roughness to accurately predict the bottom boundary layer [1]. Similarly, measurements of seafloor sediment transfer also depend of the size of the bottom boundary layer and hence, seafloor roughness [2]. Acoustic remote sensing methods used to characterize the seafloor for habitat and sediment composition depend in part of the roughness of the seafloor at the carrier wavelength of the ensonifying signal [3]. This last application, namely, the characterization of the seafloor by remote acoustic methods, has led to consideration of methods for measuring the microbathymetry of the seafloor.

Multibeam echosounders, used throughout the world for the routine collection of bathymetric data, commonly also collect co-registered seafloor acoustic backscatter. Acoustic backscatter of the seafloor may be used to characterize the sediment type [4], the presence of gas, and the likely habitat of many benthic organisms and bottom dwelling fish [5]. However the process of interpreting seafloor backscatter is complicated by the fact that a large portion of the returned signal at non-normal angles is dependent on the seafloor

roughness at a scale as small as the wavelength of the acoustic carrier frequency [3]. A shallow water system operating at 200 kHz has a nominal wavelength of just 0.75 cm. This scale is far smaller than that resolvable by the sonar's own bathymetric measurements. Therefore variations in seafloor backscatter may be recorded due to unknown changes in roughness with no change in sediment composition or other factors. Moreover because seafloor roughness may not be isotropic, multiple measures of seafloor backscatter measured on different headings (and hence ensonifying angles) over the same seafloor may produce different results. This paper presents preliminary results of the use of a GO-PRO underwater stereo camera system to measure the roughness of the seafloor at scales comparable to those that affect acoustic backscatter from commonly used bathymetric sonar systems. Section II describes the cameras and their operation, Section III describes the algorithms used to create dense 3D point clouds from pairs of stereo images and Section IV provides some preliminary results captured thus far.

## II. THE STEREO CAMERA KIT

A GO-PRO camera is a fixed focus, 11 mega-pixel, still-frame (or 1080p high-definition video) camera measuring just 42 mm x 60 mm x 30 mm (Fig. 1). Although the camera has no viewfinder or LCD screen, its small form factor combined with a standard waterproof enclosure has made it popular with sports enthusiasts for underwater and extreme sports footage. (The camera is frequently mounted to the chest or head while skiing, surfing, sky diving etc.) A stereo camera kit is available for the GO-PRO camera, which provides a waterproof enclosure (to 200 ft) for a pair of cameras connected by a synchronization cable for synchronized video or still photography. When connected, pairs of cameras take synchronized still photos and an automatic timer setting allows the taking of still images at regular intervals (the interval is user selectable from 2 to 60 sec) without user interaction. The cameras and kit provided a low cost and easily used system for capturing stereo image pairs for micro-bathymetry.

## III. CREATING 3D POINT CLOUDS FROM STEREO CAMERA PAIRS

Stereo cameras take synchronized images of a scene from differing vantage points. In general, by knowing the translation and rotation of one camera relative to the second, one can



Fig. 1. GOPRO Cameras with stereo housing and synchronization cable.

establish the location of an object in 3D space when the object is uniquely identified in both images. Conjugate pixels in each image provide pointing vectors from each camera's focal point whose intersection locates the object in question. Challenges to stereo camera reconstruction involve methods to match pixels corresponding to common objects in both images.

Methods of producing 3D point clouds from stereo camera pairs in air are well established [6]. However methods for underwater photography are relatively nascent, in part, due to the complicating factors related to the light-propagating medium. Underwater images often appear blurry due to the scattering effects of water molecules and particles in the water column. Moreover, ambient lighting can be irregular and the displacement between the cameras enhances the effect often resulting in different exposures between images. For these reasons, color is, in general, an unreliable metric for matching pixels between image pairs. Instead images are converted to gray scale and the resulting image texture proves more reliable.

The process of creating 3D point clouds from stereo camera pairs first involves calibration of the camera pair to measure the distortion of optical system of each camera (intrinsic parameters) and to establish the translational rotational offsets and between their optical axes (extrinsic parameters). The calibration has been done with the Camera Calibration Toolbox for MATLAB [7]. Images of a checkerboard pattern were taken at several orientations. Calibration images are taken underwater for accurate compensation for lens distortion on subsequent underwater imagery.

A pair of images selected for processing are first cropped and resampled to a smaller size for convenience. GO-PRO cameras have a field of view of 170 degrees in air. Such a wide field of view imparts severe distortion to portions of the image near the edges, which are difficult to capture in the calibration process. One eighth of the image is removed from each side, reducing a 3840 x 2880 pixel image to 2880 x 2160 pixels.

Lens related distortion is then compensated for, and images are rectified using standard methods [8]. Rectification produces two images with epipoles at infinity, such that pixels in each

horizontal row fall along an epipolar line. Searches for conjugate pixels between images may then be simplified to a search primarily in one dimension. The distance of a matching pixel in one image relative to the other along each row of rectified images is termed the horizontal disparity which is directly related to the range from the stereo rig to an imaged scene.

Rectified images are next resized to 720 x 540 through an averaging process. This step is optional but was undertaken in the preliminary tests to reduce the processing time. The step involves averaging neighboring pixels rather than simple decimation. The distinction is important as decimation aliases high frequency texture components reducing the effectiveness of attempts to match pixels between images. (Processing of full images is possible and will likely result in even a denser point cloud.)

The SIFT algorithm [9] is next used to determine the range of likely disparities between the images and to provide seeds for the subsequent dense matching algorithm. The SIFT algorithm creates a Laplacian pyramid of images and utilizes the difference of Laplacians to extract points of interest at different spatial scales with subpixel accuracy. Extracted points are then matched across the images using the similarity of descriptors associated with each point to produce a sparse set of matching points. To meet the high-resolution requirements of this project dense matching is required. Therefore, matches are, when possible, found for all pixels in the image in subsequent steps.

The methodology described above has been developed in the framework of the project with NOAA South-West Fisheries Science Center with the aim of detection and measurement of live fishes [10]. In the current work this research has been extended by segmenting the rectified images based on their texture [11,12]. The rationale of image segmentation for stereo processing is that surfaces, which are smoothly varying in 3D space (and hence in disparity space), are likely to have homogeneous texture and thus appear in the same segment. Also, a sudden change in disparity (due to an occlusion, for example) usually manifests itself as textural or colorimetric change and thus cause a boundary between neighboring segments. Segmentation is performed at different levels of granularity, and a level with 600-800 segments is selected for further processing. In this case, each segment has area of approximately 500 pixels – a sufficient amount to collect representative histograms and small enough to guarantee an absence of disparity jumps within a single segment.

Each segment is considered separately. For each pixel  $P_0$  in a segment a number of potential candidates for its conjugate  $P_0'$  in the other image are chosen. A window of pixels in the vicinity of  $P_0$  and  $P_0'$  and within the same segment are selected. The maximum number of pixels in a window is 7 x 7, but they are arranged in variable patterns. The smallest scale consists of a 7 x 7 window, the next scale a 13\*13 window, and the last, the 5-th scale a 31 x 31 window. Thus, with the same calculation complexity the similarity between regions can be detected on a variety of scales. Normalize cross correlation (NCC) scores are calculated for each window. Locations of  $P_0'$

with the highest NCC scores are recorded. Pixels with conjugates detected by SIFT matching are considered to have the highest NCC score of 1, but other potential candidates are found for them too.

Next, histograms of disparities for each individual segment are constructed. Two histograms are created: one consists of only top-scoring candidates, while the second contains all recorded candidates for all pixels in the segment. In most cases disparities for incorrect matches are distributed uniformly within the search range, while correct disparities are localized in a narrow interval, so that each histogram has a distinct peak. When the texture in the image is rich and distinct, the histogram of top scorers usually contains a single peak corresponding to the correct match for the segment as a whole. However when the texture is not well pronounced the histogram of top scorers is less reliable. In this case a histogram of all recorded candidates better represent the disparity of the segment. Both histograms are processed and the dominant peaks are determined. In the case of a single, co-located peak in both histograms, it is accepted as a final solution for a segment. In cases of several peaks with comparable dominance each associated solution is investigated individually.

To resolve cases of multiple histogram peaks showing no agreement to the correct disparity the candidate values are compared to the seed values (from SIFT matching). First disparities corresponding to the seed values and those corresponding to the peak maximum are set. In an iterative process, the neighboring disparities for resolved disparities are chosen such that the difference with already set values are minimized, resulting in the smoothest possible solution. As a check of these results two quantitative characteristics are considered: average NCC score of chosen disparities and average roughness of disparities within a segment, where

roughness is defined as the sum of absolute differences between neighbor disparities divided by number of neighboring pairs of pixels. Practice shows that the correct solution does not necessarily have the highest NCC score and the lowest roughness, so discarding all but the best solution might lead to a wrong result. Hence if a few top solutions minimizing the transition from seed values have comparable scores they are all kept to make the final decision at a final stage.

When histograms of disparities within a segment, comparison with SIFT determined seeds, amplitude of NCC scores and local roughness of NCC scores all fail to definitively resolve the correct disparity, the candidates are compared to those in adjacent segments. Disparity ranges in successful segments neighboring the one with several equally good solutions are compared. Again the assumption of local smoothness is utilized. The investigated segment is likely to have smooth transition of disparities with the majority of its neighbors. If this condition is not fulfilled for any of the kept solutions the segment is considered an outlier and invalidated (its pixels are not used in triangulation).

Methods described thus far determine the disparity between matching pixels in the two images with the resolution of a single pixel. To gain subpixel resolution, a parabola is fit to the disparities of neighboring pixels centered on the pixel of interest and the peak of the paraboloid is chosen for the final disparity measure.

The camera calibration along with the calculated disparities between matching pixels are used to calculate the location of each object in 3D space by triangulating the intersection which originate at each camera's focal point and whose direction is determined by their disparity using standard methods [6]. These results produce a point cloud, which is used to generate the surfaces presented in the next section.

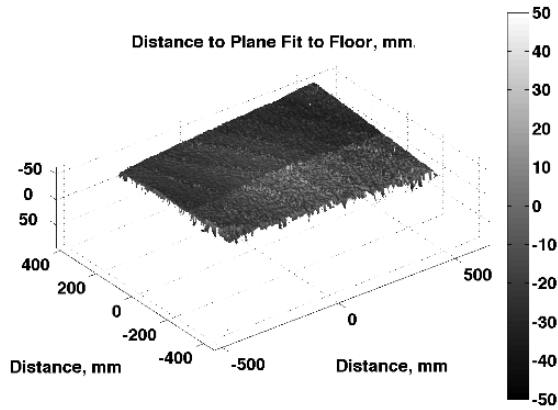
#### IV. RESULTS

Preliminary tests were conducted in air to test the setup and method. Figure 1 shows the left camera image of an office carpet and blue towel laid flat on the surface. Figure 2a shows a 3 mm x 3 mm median grid of the point cloud data after subtracting the point heights from a plane fit to the portion of the data associated with the carpet. The RMS deviation of the data to the plane is 9 mm. Careful examination of the surface shows two artifacts. The first is a slight curvature to the surface revealed as a lightening of the gray-scale shaded image in the center of the grid. This curvature results from an imperfect correction for lens distortion and was left uncorrected for to illustrate the effect. The second is a small 5 mm irregularity in the surface creating bands in the gray-scale height. This results from an inability of the sub-pixel disparity algorithm to discriminate disparities at the sub-pixel level for the textures provided by these surfaces.

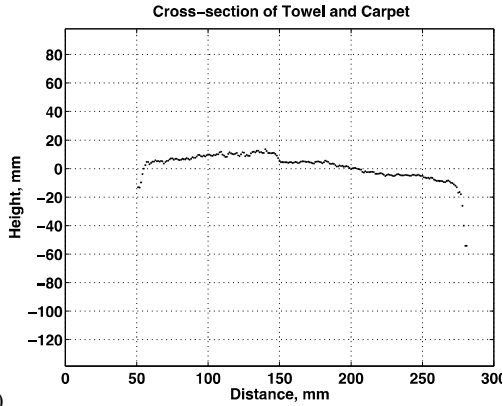
Image of Towel and Carpet.



Fig. 2. One of a pair of stereo camera images used for initial in-air testing of the system's ability to generate dense 3D point clouds of textured surfaces.



a)



b)

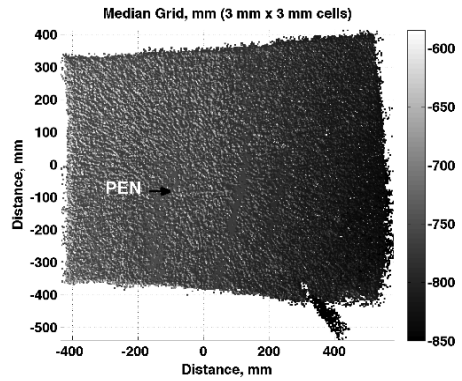
Fig. 4. The surface in a) is a grid of the residuals to a plane fit to the 3D point cloud calculated from stereo reconstruction of the towel and carpet scene shown in Figure 2. A cross-section of the surface is shown in b) showing the towel's 7 mm height relative to the floor. Residual curvature in the surface results from an imperfect camera calibration, left uncorrected to illustrate the effect.

Figure 2b shows a cross-section of the data set in which the towel is clearly visible as a 7 mm increase in surface height at a distance of 150 mm from the edge of the plot. While errors in our calibration methods left residual curvature to the surface these results were sufficiently promising to continue investigation of the method.

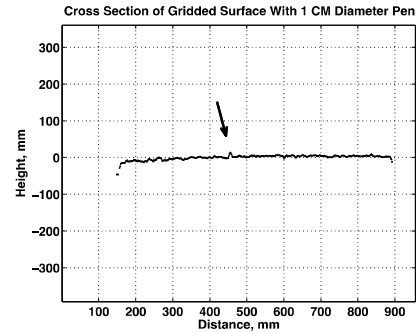
Figure 4 shows the left image for a second test photo in air, in which a pen was imaged on a wet sandy beach. The grain size and texture of the beach provided a means to test the algorithms in a real-world scenario and to adjust the methods to obtain the best image for stereo reconstruction. Here the pen is clearly recognizable in the surface plot and cross-section provided in Figures 5a and 5b respectively.



Fig. 5. One of a pair of stereo camera images used for initial in-air testing of the system's ability to generate dense 3D point clouds. Here a pen was laid on a sandy beach to further test the resolving capability of the method.



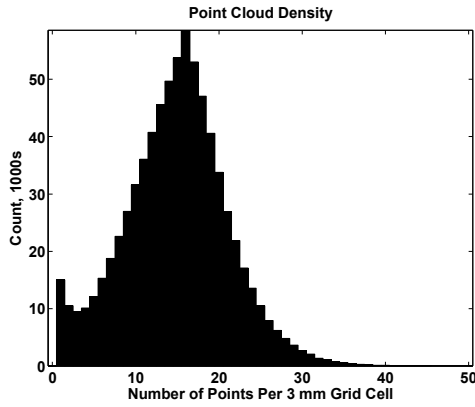
a)



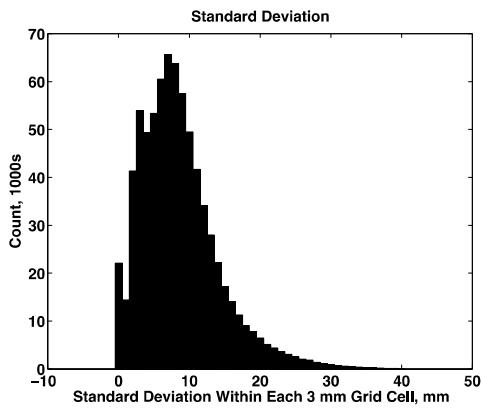
b)

Fig. 3. The surface in a) is a 3 mm grid of the 3D point cloud calculated from stereo reconstruction of the beach and pen scene in Fig. 4. A cross-section of the surface is shown in b) where the pen is clearly evident as a 1 cm bump on the surface.





a)



b)

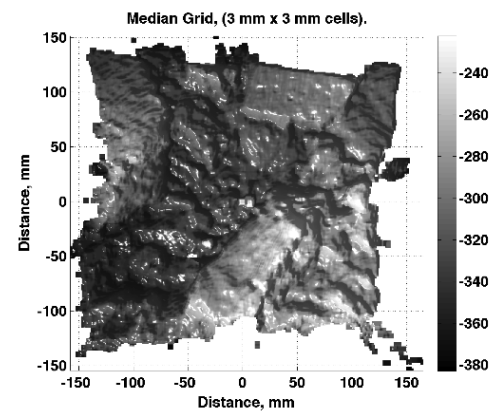
Fig. 7. Here the number of points per 3 mm grid node and their standard deviation are plotted for the surface shown in Fig. 4a. The images used for this reconstruction were taken at approximately 0.7 m above the surface and were decimated by a factor of 4 prior to stereo reconstruction.

Figures 6a and 6b show the number of data points per grid node and their standard deviation respectively for this image. The image was taken approximately 0.7 m from the surface, which is commensurate with the altitude from which we expect to take images in subsequent field experiments, allowing for adequate light and water clarity. While the surfaces generated from stereo image pairs can be noisy, dense matching of all available pixels generally provides sufficient data density to remove much of the noise in subsequent averaging.

Figures 7a and 7b show an underwater test image and the resulting surface generated from the stereo reconstruction. The cameras were recalibrated underwater for this test and although no man-made structures exist in this image to provide a measure that the scene is generated correctly, a qualitative analysis is possible. Major features (large stones and cobbles) in the scene are well represented. Segments that fall in a shadow



a)



b)

Fig. 6. After underwater calibration of the stereo system, an underwater scene was taken shown in a) above. A 3 mm grid of the resulting point cloud is shown in b).

(and therefore have not texture) or are occluded from the view of either camera produce outliers. These are shown as small white patches in this vertical view of the surface as the outliers are invariably shallow with respect to the surface and off the color scale. Objects on the order of 1 cm are resolved, although spaces between objects are often smoothed, in part due to the gridding algorithm.

## V. CONCLUSION

A GO-PRO underwater camera stereo camera rig has been used to generate 3 mm resolution grids of seafloor bathymetry for the purposes of seafloor characterization. Measurements at this scale allow characterization of the micro-roughness of an area, which is critical in the modeling of many processes including laminar and turbulent flow, seafloor sediment transport and acoustic backscatter.

Images taken less than 1 m from the surface with an orientation nearly normal were found to provide adequate resolution and uniform density of the resulting point cloud. Underwater stereo imagery is generally more challenging than that in air, as light emanating from a point on the seafloor is scattered by the water column producing blurring or hazing effect that complicates matching of pixels between images. This blurring effect combined with homogeneous fine grain sediments (silt and mud) requires images as close as 20-30 cm for the cameras to resolve individual grains for matching. Large objects imaged from a stereo camera pair produce typically produce rich texture, but complicate processing with occlusions (portions of a scene viewable in one image but not in the other) which produce outliers.

The GO-PRO underwater camera stereo system is not found to be ideal, having a quite short baseline (3.5 cm), which limits resolution of the resulting point cloud. When the scene has poor texture the sub-pixel algorithm used here can fail to improve the resolution of the system beyond that produced by the baseline alone. The final point cloud may appear to have discrete steps corresponding to integer pixel disparities as a result. Moreover, the cameras require high light levels to take adequately illuminated images. At deeper depths artificial lighting may be required. None-the-less, great convenience is found in the relatively low cost, prepackaged system and measures of seafloor micro-bathymetry capable of resolving seafloor features less than 1 cm in amplitude were possible.

## VI. ACKNOWLEDGEMENT

This work was funded by NOAA Grants NA05NOS4001153 and NA10NOS4000073

## VII. REFERENCES

- [1] R. G. Dean and R. A. Dalrymple, *Water wave mechanics for engineers and scientists*. World Scientific, 1991.
- [2] R. G. Dean and R. A. Dalrymple, *Coastal Processes with Engineering Applications*. Cambridge University Press, 2004.
- [3] X. Lurton, *An Introduction to Underwater Acoustics: Principles and Applications*, 2nd ed. Springer, 2010.
- [4] Fonseca and L. Mayer, "Remote estimation of surficial seafloor properties through the application Angular Range Analysis to multibeam sonar data," *Mar Geophys Res*, vol. 28, no. 2, pp. 119–126, Jun. 2007.
- [5] P. T. Harris and E. K. Baker, *Seafloor Geomorphology as Benthic Habitat: GeoHAB Atlas of Seafloor Geomorphic Features and Benthic Habitats*. Elsevier, 2011.
- [6] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge, UK; New York: Cambridge University Press, 2003.
- [7] J.-Y. Bouguet, *Camera Calibration Toolbox for MATLAB*, [http://www.vision.caltech.edu/bouguetj/calib\\_doc/](http://www.vision.caltech.edu/bouguetj/calib_doc/).
- [8] A. Fusiello, E. Trucco, and A. Verri, "A compact algorithm for rectification of stereo pairs," *Machine Vision and Applications*, vol. 12, no. 1, pp. 16–22, 2000.
- [9] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [10] Y. Rzhhanov and G. Cutter, "StereoMeasure and StereoFeatures: measuring fish and reconstructing scenes using underwater stereo photographs," National Marine Fisheries Service, Seattle, Washington, NOAA Technical Memorandum NMFS-F/SPO-121, Sep. 2010.
- [11] Y. Deng and B. S. Manjunath, "Unsupervised segmentation of color-texture regions in images and video," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 8, pp. 800–810, Aug. 2001.
- [12] L. Prasad and A. N. Skourikhine, "Vectorized image segmentation via trixel agglomeration," *Pattern Recognition*, vol. 39, no. 4, pp. 501–514, Apr. 2006.