GoPro Modeling and Application in Opti-Acoustic Stereo Imaging*

Shahriar Negahdaripour[2], Hitesh Kyatham[1], Michael Xu[1], Xiaomin Lin[1], Yiannis Aloimonos[1], Miao Yu[1]

[1]*Maryland Robotics Center, University of Maryland*

[2]*ECE Department, University of Miami, Coral Gables, FL*

***Abstract*—This paper deals with a number of computer vision techniques for the integration and interpretation of visual cues in RGB and forward-look sonar images. To utilize a low-cost GoPro stereo imaging system with dedicated water-proof housing and an Oculus sonar, we perform camera calibration by ray tracing to account for the impact of refraction at the housing glass ports, and calculate the relative poses of all three imaging systems. Utilizing the data for our calibrated system, we describe and assess certain 3-D reconstruction methods to determine the relative position of various scene targets for collision avoidance, to generate 3-D object models, and to enhance RGB images by haze removal. Experiments with the real images of various targets in a pool and a water tank under both good visibility and turbidity are presented to demonstrate some advantages in the integration of multi-modal visual cues. Collectively, these methods are targeted for the realization of capabilities that enhance marine robotics perception and autonomy in near-seabed operations.**

Key words: Forward-look (FL) Sonar; Opti-Acoustic Stereo Imaging; Oculus Sonar; Didson; GoPro RGB Camera; 3-D Reconstruction.

I. Introduction

Underwater robot perception plays a crucial role in a range of underwater scientific and commercial explorations; archaeology [38]- [7], coral reef management [13], [23], and offshore oil industry operations [40]- [41]. Precision landmark-based localization, navigation, and reacquisition as well as 3-D reconstruction, modeling and mapping are some robot capabilities that require the recovery of quantitative 3-D scene information from 2-D optical images in relatively clear waters. Alternatively, 3-D reconstruction methods based solely on sonar data may be applied in turbid waters; e.g., [1]- [5], [15], [17], [19], [24], [34]- [36]. More importantly, key advantages are achieved in multi-modal opti-acoustic stereo imaging by the integration of optical and sonar visual cues, where the extraction of dense prominent features is hindered by increased turbidity, but is still feasible to identify a handful of sparse features, locate structural features (edges), and (or) detect occluding contours; e.g., [10], [25]- [28].

In deploying RGB cameras for monocular and stereo vision, the internal calibration establishes the relationship between image measurements and various 3-D scene properties. Moreover, where available, measurements from a 2-D forward-look sonar with known pose relative to the RGB camera(s) can be integrated for improved robustness and accuracy. For underwater deployment, some optical cameras utilize a waterproof housing with a *flat transparent glass port*. This leads to the deviation from the ideal perspective projection model; nonlinear refraction of optical rays at the interface occur due to variation of light speed through various media, namely, the air, glass and water [6], [11], [14], [22], [31], [32]. Here, the rays bend towards the surface normal, when entering a denser medium at the interface. Equivalently, the pin-hole camera model with a single projection center (SPC) – where optical rays from the scene to the camera intersect – is invalidated. To address, non-SPC ray tracing methods have been proposed for the calibration of monocular, stereo, and projector-camera system in support of structured light methods; e.g., [33].

In this work, some RGB data come from a single Sony camera in a glass dome, with no/minimal diffraction. Here, we have successfully applied the standard SPC model for intrinsic calibration. However, we also utilize data recorded with two GoPro cameras in stereo configuration, each enclosed in a water-proof housing with a flat glass port. Thus, the scene-to-image projection deviates from the SPC model; see Fig. 1(b) for a sample stereo pair of a grid of metal reflectors, also employed for opti-acoustic stereo calibration [26], [37]. Here, we apply a ray tracing scheme for intrinsic calibration, traditional extrinsic calibration for the GoPro stereo system, and opti-acoustic calibration [26] to determine the pose relative to an Oculus sonar [42]. These allow us to exploit the visual cues in optical and sonar data; see Fig. 1 (b). The data from the calibrated optical and opti-acoustic stereo imaging system are utilized to explore the application and to assess performance in selected 3-D reconstruction techniques.

It is noted that this paper covers several topics incorporated in an elective graduate course (*Underwater Robot Perception*), in the Maryland Applied Graduate Engineering (MAGE) program [43] [1]. Moreover, experiments are presented for 3-D scene reconstruction from overlapping GoPro and Oculus images, similar to those carried out in regular and term

This work was supported by USDA NIFA sustainable agriculture system program under award number 20206801231805.

[1] This elective course was designed and offered in Sring'24 semester during a sabbatical leave of the first author, as a Visiting Professor in the Department of Mechanical Engineering, University of Maryland, College Park, MD.

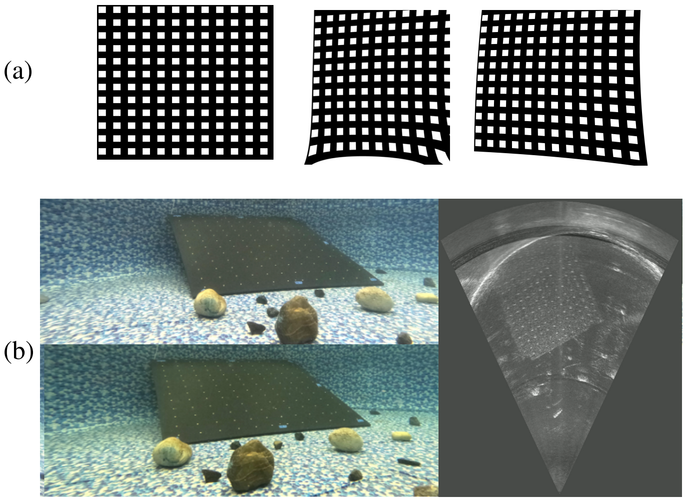

Fig. 1. (a): Rectifying perfect grid in GoPro dat yielding distorted left and right image grids; (b) sample GoPro stereo pair (pair 3 in Fig 2) and Oculus sonar image used in opti-acoustic calibration.

projects with images chosen by individual students. These compare the accuracy in 3-D terrain and target reconstruction by single-/dual-modality stereo imaging, and explore how and if the single-image haze removal by dark channel prior [16] may be improved by exploiting these solutions, namely, to calculate the transmission field and airlight more effectively. Our results and analysis sheds light on various scenarios where the methods presented here may improve autonomy in the operation of underwater robots.

In the remaining sections, technical background on GoPro calibration and multi-view reconstruction is presented in section II. Experimental results for various methods are given in section III. The summary, conclusions and future efforts are provided in Section IV.

II. Technical Background

A. Calibration

Deployment of cameras for underwater scene/object reconstruction calls for the treatment of refraction at various air, glass and water interfaces, e.g., [32]. The so-called refractive structure-from-motion techniques directly model the impact on the epipolar geometry in two views; e.g., [12], [20], [21], [29]. However, general applications calls for the calibration of 3D-to-2D projection geometry by ray tracing [6], [11], [14], [22], [31]. These include but is not limited to 3-D reconstruction by multi-modal stereo imaging with a forward-look sonar (having overlapping view), structured light stereo [33], and the likes.

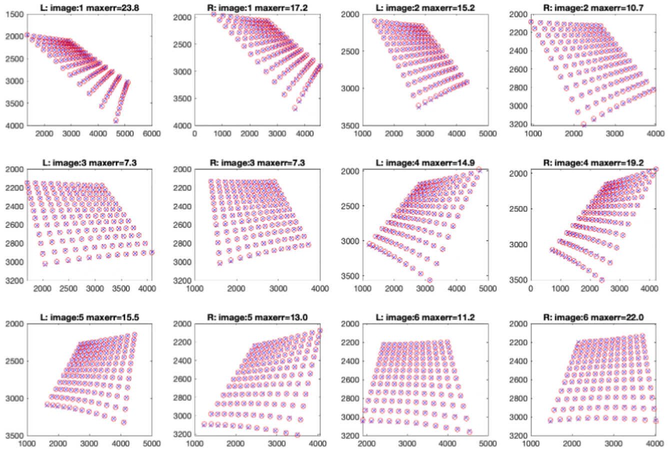

Fig. 2. Data (red circles) and reprojected calibration grid points (blue crosses) based on calibration of two GoPro cameras in stereo configuration with maximum discrepancy of 23.8 pixels.

We follow the commonly adopted approach where each pixel is assigned an incident ray direction in 3-D space. While an axial camera model with two air-glass and glass-water interfaces has been assumed [22], our experimental results confirm that simplified computations by ignoring the relatively thin glass layer of about 2.5 [mm] has negligible impact on the calibration accuracy. Accordingly, we express each optical ray by the simplified model [21], [33].

$$\hat{\mathbf{P}}_o = (\cos\gamma' - 1/t\cos\gamma)\,\hat{\mathbf{n}} + 1/t\;\hat{\mathbf{P}}_i \tag{1}$$

where γ and γ' are the angles of the so-called *entering* (air side) and *existing* (water side) rays at the interface, t is the relative water-to-air refraction index, $\hat{\mathbf{n}}$ is the optical axis (perpendicular to the interface) , and $\hat{\mathbf{P}}_i$ and $\hat{\mathbf{P}}_o = \mathbf{P}_o/|\hat{\mathbf{P}}_o|$ are unit vectors along the entering and existing ray directions. The entering ray direction $\mathbf{P}_i = (x_d/f_x, y_d/f_y, 1)$ is expressed in terms of the displaced image position $\mathbf{x}_d = (x_d, y_d)$ (by lens distortion), normalized by the focal lengths f_x and f_y in pixel units in the x and y directions, respectively. Alternatively, we utilize the ideal image positions $\mathbf{x} = (x, y)$ after lens distortion correction, according to the distortion model:

$$\begin{aligned}\mathbf{x}_d = (1 + k_1 r^2 + k_2 r^4)\mathbf{x} + &\big(2p_1\,xy + p_2(r^2 + 2x^2)\\ &\;2p_1\,xy + p_2(r^2 + 2y^2)\big)\,;\\ r^2 = \left(\tfrac{x - x_c}{f_x}\right)^2 + &\left(\tfrac{y - y_c}{f_y}\right)^2\end{aligned} \tag{2}$$

Finally, a 3-D point $\mathbf{P}_r$ projecting onto an image point $\mathbf{x}$ can be parameterized by the distance β along the existing ray:

$$\mathbf{P}_r = d\,\mathbf{P}_i + \beta\,\hat{\mathbf{P}}_o \tag{3}$$

where the distance d from the lens to the glass port is determined by calibration.

We utilize the same planar grid of acoustic reflectors at a number of distinct poses for calibrating the GoPro and opti-acoustic stereo imaging systems; see Fig. 1(b).

B. Multi-View Reconstruction

Traditional reconstruction with a calibrated stereo system involves determining depth Z by triangulation using the projection rays of correspondences $\mathbf{x}_l = (x_l, y_l)$ and $\mathbf{x}_r = (x_r, y_r)$ in the left and right images. Applying (3) to the rectfied stereo configuration leads to a closed-form least-squares solution for the distances $\{\beta_l, \beta_r\}$ along the exiting rays $\{\hat{\mathbf{P}}_{ol}, \hat{\mathbf{P}}_{or}\}$:

$$\begin{pmatrix} \hat{\mathbf{P}}_{ol}^T\hat{\mathbf{P}}_{ol} & \hat{\mathbf{P}}_{ol}^T\hat{\mathbf{P}}_{or} \\ \hat{\mathbf{P}}_{or}^T\hat{\mathbf{P}}_{ol} & \hat{\mathbf{P}}_{or}^T\hat{\mathbf{P}}_{or} \end{pmatrix} \begin{pmatrix} \beta_l \\ \beta_r \end{pmatrix} = \begin{pmatrix} d_l(\hat{\mathbf{P}}_{ol}^T(\mathbf{P}_{il} - \mathbf{P}_{ir}) + \hat{\mathbf{P}}_{ol}^T\mathbf{t} \\ -d_r(\hat{\mathbf{P}}_{or}^T(\mathbf{P}_{il} - \mathbf{P}_{ir}) - \hat{\mathbf{P}}_{or}^T\mathbf{t} \end{pmatrix} \tag{4}$$

where $\mathbf{t}$ is the baseline of the rectified stereo images.

For 3-D reconstruction from multi-modal opti-acoustic data, we note that the sonar image position $\mathbf{x}_s = \Re(\sin\theta, \cos\theta)$ is expressed in terms of the range $\Re$ and azimuth θ measurements, namely, two of the three spherical coordinates $(\Re, \theta, \phi)$ of a 3-D point $\mathbf{P}_s = \Re(\sin\theta\cos\phi, \cos\theta\cos\phi, \sin\phi)$ in the sonar coordinate system. The 3-D point is located on a circular arc defined by the beam in the azimuthal direction θ at distance $\Re$ from the sonar. Moreover, only the segment within the narrow vertical beam width $|\phi| \leq \Phi_{\text{max}}$ is relevant ; $\Phi_{\text{max}} = [3^\circ - 7^\circ]$ for most existing high-frequency forward-look imaging systems.

In a calibrated opti-acoustic system, the 3-D point $\mathbf{P}_s$ in the sonar coordinate system can be expressed in terms of the rotation matrix $\mathbf{R}_{os} = (\mathbf{r}_1; \mathbf{r}_2; \mathbf{r}_3)$ (with rows $\mathbf{r}_i$) and translation $\mathbf{t}_{os} = (t_x, t_y, t_z)^T$ of the Oculus sonar coordinate system:

$$\mathbf{P}_s = \mathbf{R}_{os}\mathbf{P}_r + \mathbf{t}_{os} \tag{5}$$

Given the opti-acoustic correspondences $\mathbf{x}_s = \Re(\sin\theta, \cos\theta)$ and $\mathbf{x} = (x, y)$, a simplified solution for the depth Z of a 3-D point in the optical coordinate system is the smaller/positive of the two intersections of the optical ray with the sonar range sphere:

$$(\mathbf{P}_r \cdot \mathbf{P}_r)Z^2 + 2(\mathbf{t}_{os}^T\mathbf{R}_{os}\mathbf{P}_r)Z - (\Re^2 - \mathbf{t}_{os} \cdot \mathbf{t}_{os}) = 0 \tag{6}$$

Alternatively, a unique solution is derived by the intersection with the azimuth plane:

$$Z = \frac{\tan\theta t_y - t_x}{(\mathbf{r}_1 - \tan\theta\mathbf{r}_2) \cdot \mathbf{P}_r} \tag{7}$$

In practice, we utilize the solution that minimizes the weighted reprojection errors in the two images [26].

When operating near the seabed, it is often feasible to identify a minimum of three *small* non-collinear bottom features in the RGB image, even in the presence of some haze. We may apply the epipolar constraint to identify the matching points (generally a small highlight) in the sonar image [25]. Consequently, we determine the 3-D coordinates of the corresponding scene features by opti-acoustic triangulation [26]. For a a relatively flat sea floor, these are sufficient to compute a plane model, which has several useful applications. For example, we can overcome a severe bottleneck in optical-image haze removal that requires the estimation of the transmission field and airlight, generally using the dark pixels in color channels; e.g., [16] (hereby referred to as the HST method). Unfortunately, the transmission field cannot be estimated reliably when the dark-channel assumptions are violated by some scene objects, near-field backscatter with artificial lighting in deep waters, sun flicker in shallow waters, and the likes. Using the seafloor plane model, we can reconstruct any bottom feature in 3-D; by the intersection of optical/sonar projection ray/arc with the seafloor plane. Finally, we can calculate the range/depth of the entire seafloor region, using which the optical image can be dehazed more effectively.

Alternatively, we can reconstruct the 3-D occluding contours of diffuse seafloor objects from their cast shadows on the bottom surface; e.g., [9]. Moreover, if range varies monotonically over the object surface, ensuring a one-to-one correspondence between pixels within the object image region and corresponding 3-D surface patches, the backscatter measurements can be employed to build a 3-D object model (based on the shape-from-shading paradigm) [8], [18].

III. Experiments

The calibration of the GoPro stereo imaging system yields the focal lengths, image centers, aspect ratios, lens distortion parameters, and an estimate of effective air-layer thickness within the camera housing, before the diffraction at the interface. Fig. 2 depicts the grid positions (red circles) in six stereo pairs, the reprojection of known 3-D grid points based on calibration parameters (blue crosses), and the maximum errors (discrepancy between reprojected points and the data). Here, the largest error of 23.8 pixels (in a 4872× 5568 image) represents a maximum error of less than 0.5%. Moreover, the stereo baseline of about 18 [cm] is of interest in assessing the 3-D reconstruction accuracy for the target ranges in our experiments. These are presented next to assess the reconstruction accuracy using the calibrated GoPro stereo system, as well as the opti-acoustic stereo imaging system with some of its key advantages.

Fig. 3 depicts two scenes comprising of rocks of various sizes on the bottom of a textured indoor pool. The top row includes both original and rectified (and dehazed) GoPro stereo images (using calibration results and the HST method [16]). The 3-D plot (bottom-right) shows the reconstruction of selected feature matches in the GoPro stereo images, both from the bottom surface and on various rocks.

For about 30 points on the bottom plane in the first scene, (indexed in white color), the plane fitting error in the bottom left of Fig. 3 gives a maximum distance of about 2 [cm] from the plane. Only points 2 and 18 on the bottom surface (colored in red, on both sides of the top right rock) have a plane fitting error of larger than 2 cm. The remaining 38 features (indexed in red color) correspond to points on various rocks, from about 5 [cm] to 25 [cm] in height. For example, the height (elevation above the seafloor) of point 47 (on the top middle rock) can be determined from the plane fitting error of nearly 5 [cm] .

The second error plot in the bottom left of Fig. 3 is the 3-D reconstruction error, namely the distance between the left and right projection rays at the estimated 3-D points; based on computed β_l and β_r from (4). Roughly 10% of the points have the largest error of just over 2 [cm] (mainly in Y direction),

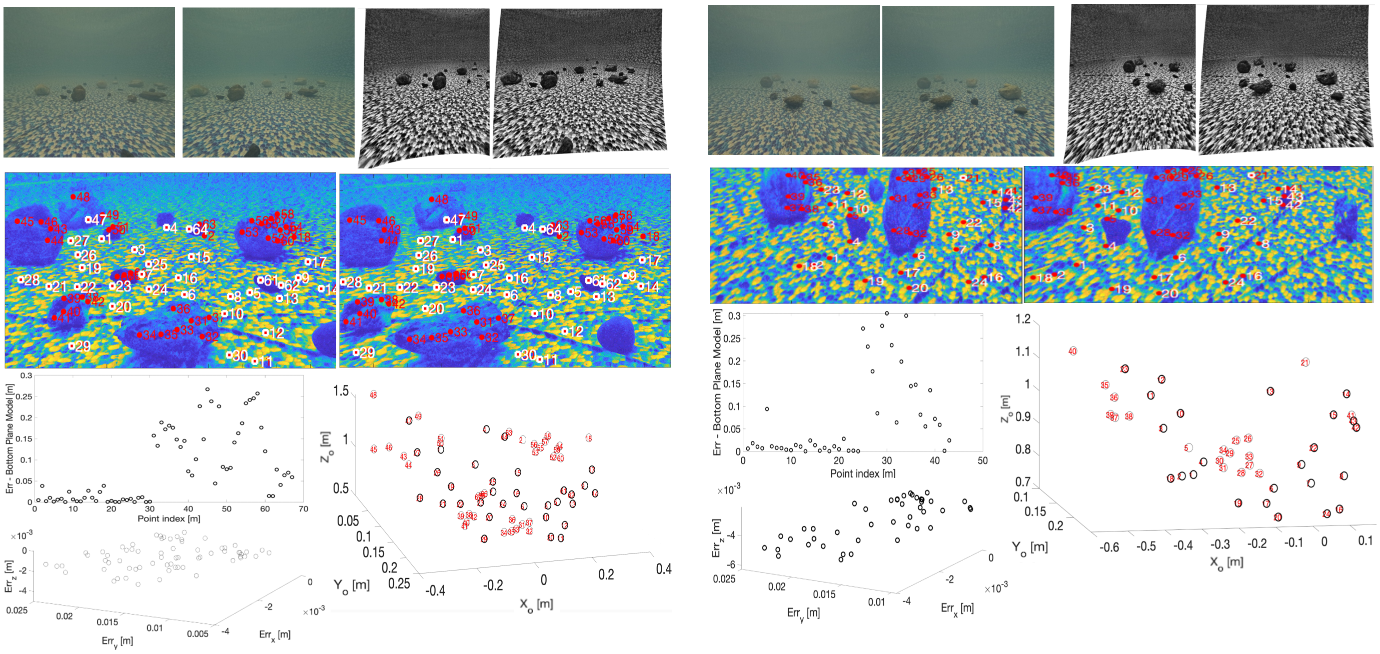

Fig. 3. Two different scenes with original GoPro stereo images, the rectified and dehazed images, and certain feature matches in rectified GoPro stereo data with their 3-D reconstructions. Bottom left are the plane fitting error for all features, and the X, Y and Z components of 3-D reconstruction errors.

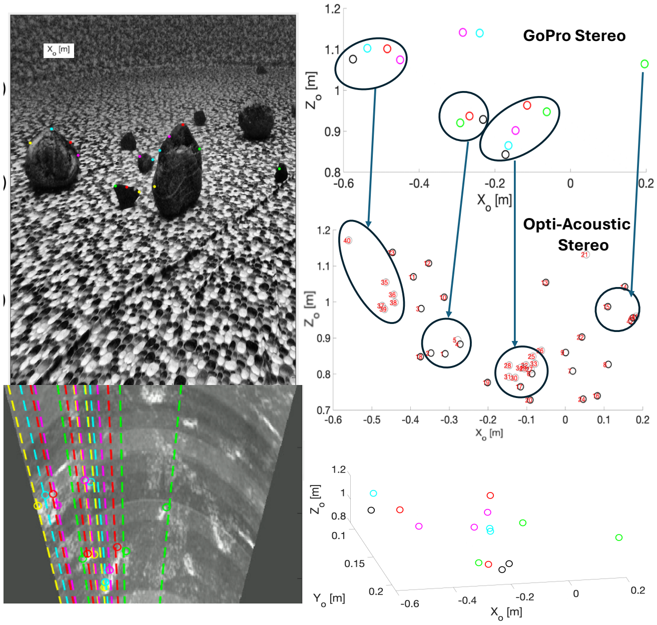

Fig. 4. Second scene in previous figure with corresponding Oculus sonar image, and selected opti-acoustic matches used for 3-D reconstruction, to compare with GoPro Stereo reconstruction from previous figure. Good correspondence is noted from XZ views.

in part due to the localization errors of the left and right matches, and in part for the imperfect calibration. Similar results are obtained for the second scene, with only 24 bottom features from a total of 43 matched features. As an example, the estimated height of 30 [cm] (above the bottom) for two points at nearly the top of the tallest rock (30 and 34) closely matches the manual measurements on the object size.

In Fig. 4, the reconstruction is performed with a few opti-acoustic correspondences. All the selected GoPro image features lie on or near the occluding contour, thus matched with the point at the intersection of the corresponding eipolar line and the top boundary of the object highlight in the sonar image. When the epipolar curve intersects multiple blobs, the correct match can generally be identified by knowing the approximate target range; e.g., by utilizing the plane equation, i.e., assuming that the feature lies on the bottom surface. Ambiguities in matching arise primarily when the other blob(s) lie within a very short distance from the true corresponding object. The clusters of reconstructed 3-D points on 4 different objects (and their XZ projections) are in good agreement with the reconstructions from the GoPro stereo data. Here, exact differences have not been calculated since the features for opti-acoustic reconstruction on (or near the occluding contours) are not the same as those automatically matched in the GoPro data.

The high accuracy of reconstructed points, albeit sparse, enables obstacle detection and collision avoidance. Moreover, applying the method in [10], an accurate dense range map can be computed uisng the 3-D points both on the bottom surface and the occluding contours of various objects. This method applies an MRF-based statistical framework, where the image intensities and known range values of reconstructed points serve as observation and hidden variables, and the opti-acoustic epipolar geometry guides the inference of the MRF by refining the neighborhood pixels.

Fig. 5 is an example of a different scene, comprising of a large number of small bottom features. Here, the first 3-D view has been selected to discriminate between the bottom features

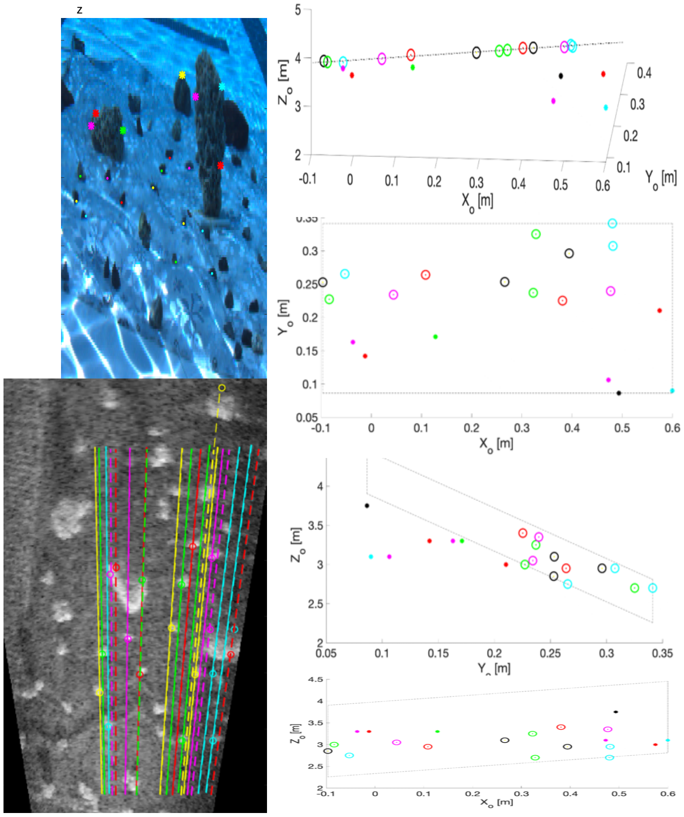

Fig. 5. Selected features in the optical image, corresponding sonar matches along epipolar curves, and 3-D reconstruction. Top 3-D plane highlights features in the ground plane, and other points on 3-D objects.

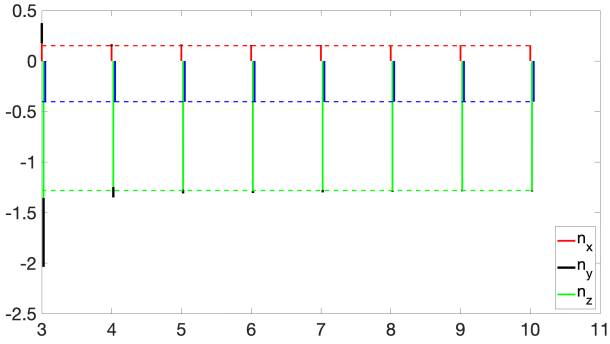

Fig. 6. (a) Estimation of bottom plane normal $\mathbf{n} = (n_x, n_y, n_z)$ using $N = 3, 4, \ldots, 10$ random bottom features, averaged over 50 samples. Color bars are the average, and black extensions show one standard deviation from the mean. Dashed lines are the estimated normal components using all bottom features. Results confirm accuracy of estimated 3-D bottom features and plane model, with little variations in the solution with as few as 5 points used to estimate plane equation.

and those on 3-D rocks. Most of the estimated 3-D object sizes are within small errors of manual measurements. For test of accuracy, we have estimated the bottom plane normal $\mathbf{n} = (n_x, n_y, n_z)$ from $N = 3, 4, \ldots, 10$ bottom features, averaged over 50 random samples. Varying N, the mean and standard deviation of the estimates are plotted in Fig. 6. Here, each color bar represents the mean estimate of one component of $\mathbf{n}$, and the black extensions show one standard deviation from the mean. It is noted that the standard deviations become negligible with as few as $N = 5$ features, confirming the high accuracy of 3-D bottom feature positions.

Referring to Fig. 7, we explore another application for the computed bottom plane model, which allows us to generate a depth map for the entire scene, depicted in (b). Despite the good visibility in the pool, some (commonly-encountered) scene characteristics are helpful in demonstrating some key deficiencies of single-image haze removal techniques that utilize dark channel prior; e.g., the HST method [16]. Here, the computations of airlight and transmission map in (d) are adversely impacted by the sun flicker (common in shallow waters) for the localization of haze-opaque pixels and the tiled black lines forming the dark pixels. This leads to the color distortion in the near field and less effective haze removal throughout the image; see (e). In contrast, the more balanced enhanced image in (c) employs the depth map in (b) for the transmission map and airlight computations.

In the next two columns of Fig. 7, we make use of the images of a tank scene, under two different turbidity levels. Here, the dark scene is illuminated non-uniformly by the deployment of a light source next to the camera. The enhanced image in (c) makes use of the estimated depth map in (b) based on the bottom plane model. For the HST method, the incorrect illumination-induced diagonal gradient in the transmission map in (d) and the white PVC cylinder as the haze-opaque region lead to the exaggerated adjustment within the central part of the image in (e) .

As our last example in utilizing the planar terrain depth map, we explore the reconstruction of three 3-D targets from backscatter cues [8], in analogy with the shape from shading paradigm [18]; see Fig. 8. This requires a one-to-one mapping from a local 3-D surface patch to an image pixel, which is achieved where the range over the surface varies monotonically in some direction. To start, the planar bottom model establishes the boundary conditions on the elevation angles at a frontal edge (comprising of points at closest range along various sonar beams) and occluding contour(s) of these objects. The two small rocks satisfy the stated condition when the long side is aligned with the sonar viewing axis. For these objects, the reconstruction closely follows each object shape. In contrast, the condition is violated for the concave blade coral, leading to a highly inaccurate 3-D object model.

IV. Summary, Conclusions and Future Efforts

This paper explores the role of RGB and forward-look sonar imaging for terrain and 3-D object modeling, enhancing visual information in RGB images, and detecting obstacles for collision-free navigation. To deploy low-cost GoPro (stereo) cameras within their water-proof housings for quantitative measurements, we have applied a ray-tracing technique for intrinsic/extrinsic stereo calibration. We have also established the pose relative to a Oculus sonar for multi-modal opti-acoustic stereo imaging.

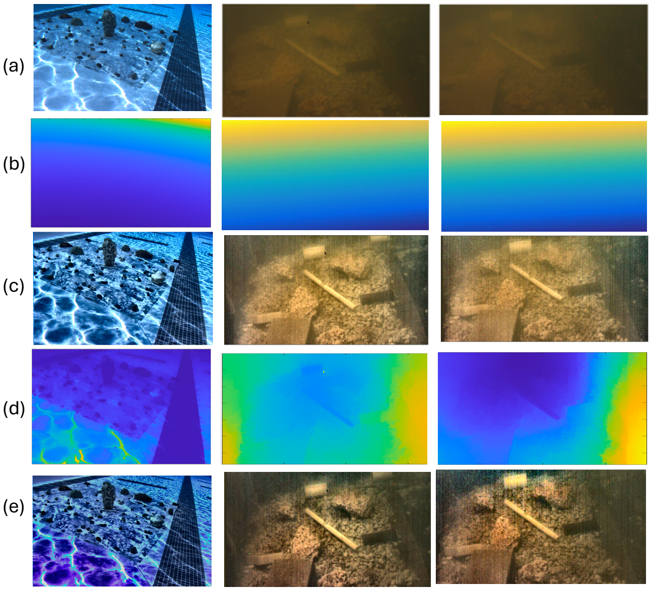

Fig. 7. (a) Original images of a pool scene (left column) and a water tank scene at two different turbidity levels (middle and right columns); (b) range map determined from bottom plane equation using 3-D positions of 3 bottom features matched in opti-acoustic stereo pairs; (c) enhanced images using estimated range map; (d) transmission filed from HST method; (e) enhanced image using dark channel (HST method [16].

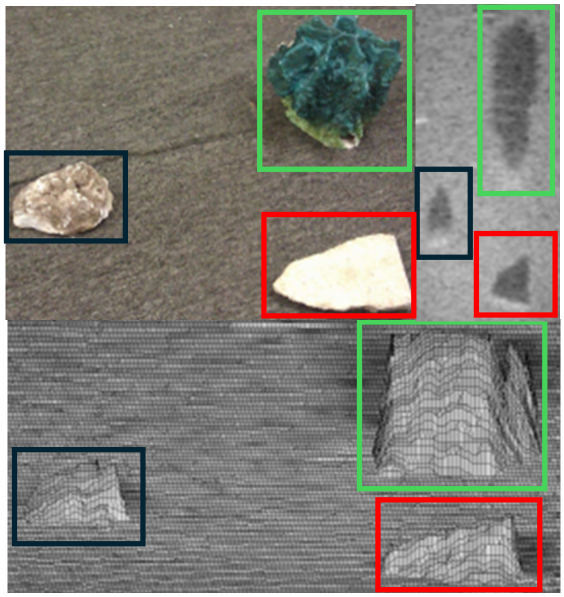

Fig. 8. Reconstruction of two small rocks, for which depth varies monotonically, and a (blue-green) concave blade coral for which multiple surface patches on each beam are located at same range, thus contributing to the measurements at the same pixel.

The Go-Pro calibration accuracy has been assessed based on the maximum reprojection error of 23.8 [pix] in a 25Mpix image (max. error of roughly 0.5%). We have also achieved a 3-D reconstruction error of no more than about 2 [cm] at average depth of about 1 [m] with a stereo baseline of 18 [cm]. The precision in opti-acoustic stereo calibration may be assessed by the estimated epipolar contours over the field of view. That is, selected object features on the occluding contour of an object in one modality yields an epipolar contour passing through the occluding contour of the matching object at roughly the same relative position. In particular, many selected scene objects (small rocks) have a relatively narrow extent horizontally, facilitating the proper assessment; i.e., with inaccurate calibration, the epipolar contour may completely miss the matching object.

Some of our data, captured with the GoPro stereo cameras and an Oculus sonar have produced consistent estimates of target distances based on both binocular and opti-acoustic stereo cues. We have also demonstrated some applications for the estimation of a flat terrain model: to employ a relatively accurate depth map over the scene to improve the transmission map computation for haze removal, and to determine 3-D object models from intensity (backscatter) measurements. For the latter, the flat bottom model establishes the boundary conditions on the elevation angles of the frontal edge and occluding contour(s) of the object resting on the bottom surface.

Underwater robot perception in attracting more attention and in particular FL sonar image processing and interpretation is become more widely researched. In this paper, we have covered roughly half of the technical contents in a first-time graduate course on marine robot perception at the University of Maryland Applied Graduate Engineering (MAGE) program, all verified and assessed through the experiments with real data. Wide adoption (fully or in part) within graduate marine robotics curricula would be instrumental in educating future researchers that are highly trained with strong technical knowledge in optical and sonar data processing.

Acknowledgement: This work is supported by the USDA NIFA Sustainable Agriculture System (SAS) Program under grant no. 2020-68012-31805. The first author acknowledges support from the MAGE program at the University of Maryland during a one-year sabbatical leave as a Visiting Professor at the Mechanical Engineering Department.

References

[1] H. Assalih et al., 3D reconstruction and motion estimation using forward looking sonar. PhD thesis, Heriot-Watt University, 2013.

[2] M. D. Aykin, and S. Negahdaripour, "Three-dimensional target reconstruction from multiple 2-D forward-scan sonar views by space carving," *IEEE J. Oceanic Engineering*, Vol 42(3), pp. 1-16, July, 2017.

[3] M. D. Aykin and S. Negahdaripour, "Forward-look 2-d sonar image formation and 3-d reconstruction," in 2013 OCEANS-San Diego, pp. 1–10, IEEE, 2013.

[4] N. Brahim, D. Guériot, S. Daniel, and B. Solaiman, "3d reconstruction of underwater scenes using didson acoustic sonar image sequences through evolutionary algorithms," *Proc. IEEE/MTS OCEANS Conf.* Spain, pp. 1–6, 2011.

[5] H. Cho, B. Kim, and S. Yu, "Auv-based underwater 3-d point cloud generation using acoustic lens-based multibeam sonar," *IEEE J. Oceanic Eng.*, vol. 43(4), pp. 856–872, 2018.

[6] A. Agrawal, S. Ramalingam, Y. Taguchi, and V. Chari, "A theory of multi-layer flat refractive geometry," *Proc. CVPR*, pp. 3346–3353, 2012.

[7] B. Allotta et al. “The ARROWS project: adapting and developing robotics technologies for underwater archaeology,” IFAC-PapersOnLine Vol 48(2), pp. 194-199, 2015.

[8] M. D. Aykin and S. Negahdaripour, "Forward-look 2-D sonar image formation and 3-D reconstruction," Proc. IEEE/MTS Oceans’13 Conference - San Diego, San Diego, CA, USA, 2013,

[9] M. D. Aykin, and S. Negahdaripour, On feature matching and image registration for two-dimensional forward-scan sonar imaging,” *J. Field Robotics*, Vol 30(4), pp. 602-623, July/August 2013

[10] M. Babaee, and S. Negahdaripour ”3-D object modeling from 2-D occluding contour correspondences by opti-acoustic stereo imaging,” Computer Vision Image Understanding, 132: 56-74, 2015.

[11] R. Li, H. Li, W. Zou, R. Smith, and T. Curran, “Quantitative photogrammetric analysis of digital underwater video imagery,” *IEEE J. Oceanic Engineering*, Vol22(2), pp. 364 –375, April, 1997.

[12] F. Chadebecq, F. Vasconcelos, G. Dwyer, R. Lacher, S. Ourselin, T. Vercauteren, D. Stoyanov, “Refractive structure-from-motion through a flat refractive interface,” *Proc. IEEE ICCV*, pp. 5315-5323, 2017.

[13] A.C.R. Gleason, D. Lirman, D. Williams, N.R. Gracias, B.E. Gintert, H. Madjidi, R.P. Reid, G.C. Boynton, S. Negahdaripour, M. Miller, P. Kramer, “Documenting hurricane impacts on coral reefs using two-dimensional video-mosaic technology,” *Marine Ecology*, Vol 28(2), pp. 254-258, June, 2007.

[14] M. D. Grossberg, and S. K. Nayar, “The raxel imaging model and ray-based calibration,” *Int. J Comp Vision*, 61(2), pp. 119–137, Feb. 2005.

[15] T. Guerneve, K. Subr, and Y. Petillot, “Three-dimensional reconstruction of underwater objects using wide-aperture imaging sonar,” *Journal of Field Robotics*, Vol 35(6), pp. 890-905, September, 2018.

[16] K. He, J. Sun, and X. Tang, “Single image haze removal using dark channel prior,” *IEEE Conf. Computer Vision Pattern Recognition*, Miami, FL, 2009

[17] Henson, B.T., and Zakharov, Y.V., “Attitude-trajectory estimation for forward-looking multibeam sonar based on acoustic image registration,” *IEEE J. Oceanic Engineering*, Vol 44(3), pp. 753 - 766, July, 2019.

[18] B. Horn and M.J. Brooks, *Shape from shading*, MIT press, 1989.

[19] T. Huang and M. Kaess, “Towards acoustic structure from motion for imaging sonar,” in Proc. IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems, IROS, (Hamburg, Germany), pp. 758–765, Sept. 2015.

[20] A. Jordt-Sedlazeck, and R. Koch, “Refractive structure-from-motion on underwater images,” *Proc. ICCV*, Sydney, NSW, Australia, pp. 57-64, December, 2013.

[21] L. Kang, L. Wu, and Y.-H. Yang. “Two-view underwater structure and motion for cameras under flat refractive interfaces,” *Proc. of ECCV*, pp. 303–316, 2012.

[22] R. Kawahara, S. Nobuhara, and T. Matsuyama, “A pixel-wise varifocal camera model for efficient forward projection and linear extrinsic calibration of underwater cameras with flat housings,” *Proc. IEEE ICCV Workshops*, Sydney, NSW, Australia, December, 2013.

[23] D. Lirman, N. Gracias, B. Gintert, A. Gleason, P.R. Reid, S. Negahdaripour, P. Kramer, ”Development and application of a video-mosaic survey technology to document the status of coral reef communities,” Environmental Monitoring and Assessment, Vol125(1-3), pp. 59-73, 2007.

[24] S. Negahdaripour, “Application of FS sonar stereo for 3-D scene reconstruction,” *IEEE . Oceanic Eng.*, Vol 45(2), pp. 547-562, April, 2020.

[25] S. Negahdaripour, “Epipolar geometry of opti-acoustic stereo imaging, *IEEE Trans. PAMI*, Vol 20(11), pp. 1776-1788, October, 2007.

[26] S. Negahdaripour, H. Sekkati, H. Pirsiavash, “Opti-acoustic stereo imaging: On system calibration and 3-D target reconstruction,” *IEEE Trans. Image Processing*, Vol 18(6), pp. 1203-1214, June, 2009.

[27] M. Qadri, K. Zhang, A. Hinduja, M. Kaess, A. Pediredla, and C. A. Metzler, “Aoneus: A neural rendering framework for acoustic-optical sensor fusion." *Proc. ACM SIGGRAPH 2024 Conference Papers*, pp. 1-12. 2024.

[28] Z. Qu, O. Vengurlekar, M. Qadri, K. Zhang, M. Kaess, C. A. Metzler, S. Jayasuriya, and A. Pediredla, “Z-Splat: Z-Axis Gaussian Splatting for Camera-Sonar Fusion,” arXiv preprint arXiv:2404.04687 (2024).

[29] M. She, F. Seegraber, D. Nakath, and K. Koser, “Refractive COLMAP: refractive structure-from-motion revisited,” *arXiv:2403.08640v1 [cs.CV]* Mar, 13, 2024.

[30] A. Shukla, and H. Karki, “Application of robotics in offshore oil and gas industry— A review Part II’,” *Robotics and Autonomous Systems*, Vol 75, Part B, pp. 508-524, January, 2016.

[31] P. Sturm, and S. Ramalingam, “A generic concept for camera calibration, ” *8th European Conference Computer Vision (ECCV ’04)*, Prague, Czech Republic. pp.1-13, May, 2004,

[32] T. Treibitz, Y. Schechner, C. Kunz and H. Singh, "Flat refractive geometry," lit IEEE T. Pattern Analysis Machine Intelligence, Vol 34(1), pp. 51-65, January, 2012.

[33] Y. Wang, S. Negahdaripour, and M. D. Aykin, “Calibration and 3D reconstruction of underwater objects with non-single-view projection model by structured light stereo imaging,” *Applied Optics*, Vol. 55(24), August, 2016

[34] Y. Wang, Y. Ji, D. Liu, H. Tsuchiya, A. Yamashita, and H. Asama, “Elevation angle estimation in 2d acoustic images using pseudo front view,” *IEEE Robotics and Automation Letters*, vol. 6(2), pp. 1535–1542, 2021.

[35] E. Westman, I. Gkioulekas, and M. Kaess, “A volumetric albedo framework for 3d imaging sonar reconstruction,” *IEEE Int. Conf. Robotics and Automation (ICRA)*, pp. 9645–9651, 2020.

[36] B. Zerr and B. Stage, “Three-dimensional reconstruction of underwater objects from a sequence of sonar images,” *Proc. 3rd 17 IEEE Int. Conference Image Processing*, vol. 3, (Lausanne, Switzerland), pp. 927–930, September, 1996.

[37] D. Yang, B. He, M. Zhu and J. Liu, "An Extrinsic Calibration Method with closed-form solution for inderwater opti-acoustic imaging system," *IEEE T. Instrumentation and Measurement*, Vol. 69(9), pp. 6828-6842, Sept. 2020

[38] https://archeologie.culture.gouv.fr/archeo-sous-marine/en/robots-are-future

[39] https://today.ucsd.edu/story/archaeologists-show-unmanned-robotic-vehicles-offer-solution-to-challenges-mapping-underwater-sites

[40] https://www.offshore-mag.com/production/article/14206250/offshore-oil-and-gas-industry-embraces-robotic-technology

[41] https://www.offshore-technology.com/features/robotics-oil-gas/

[42] https://www.blueprintsubsea.com/oculus/

[43] https://mage.umd.edu/